



Evaluation and projections of the East Asian summer monsoon in a perturbed parameter ensemble

John W. Rostron¹ · David M. H. Sexton¹ · Kalli Furtado¹ · Maria João Carvalho¹ · Sean F. Milton¹ · José M. Rodríguez¹ · Wenxia Zhang²

Received: 7 March 2022 / Accepted: 13 September 2022 / Published online: 3 November 2022
© Crown Copyright as represented by Met Office 2022 2022

Abstract

The East Asian summer monsoon (EASM) is a dominant driver of East Asian climate, with variations in its strength potentially impacting the livelihoods of millions of people. Understanding, predicting, and assessing uncertainties in these variations are therefore important area of research. Here, we present a study of the projected twenty-first century changes in the EASM using a ‘perturbed parameter ensemble’ (PPE) of HadGEM3-GC3.05 coupled climate models, which samples uncertainties arising from differences in model parameter values. We show that the performance of PPE members for leading order EASM metrics is comparable to CMIP5 and CMIP6 models in many respects. But the PPE also exposes model biases which exist for almost all parameter combinations. These ‘structural’ biases are found mainly to affect metrics for the low-level circulation. We also show that future changes in regional circulation and precipitation are projected consistently across the PPE members. A more detailed moisture budget analysis of the precipitation changes in a region covering the Yangtze River valley shows that the spread of these changes is mainly due to spread in dynamic responses. We also perform parameter sensitivity analyses and find that a parameter controlling the amplitude of deep-level entrainment is the main driver of spread in the PPE’s representation of the EASM circulation. Finally, we discuss how the information provided by the PPE may be used in practice, considering the plausibility of the models, and giving examples of ways to sub-select ensemble members to capture the diversity in the moisture budget changes.

Keywords Perturbed parameter ensembles · East Asian summer monsoon · Model evaluation · Climate projections · Sub-selection

1 Introduction

Monsoon systems are a key driver of seasonal variability throughout the tropics, directly affecting the livelihoods of over two-thirds of the world’s population (Sperber et al. 2013). Their characteristic reversal of winds in the lower troposphere, and associated variations in rainfall, are driven by seasonal variations in solar insolation, with substantial differences in local influences (e.g. land-sea temperature contrasts, orography) giving rise to distinct systems

throughout Africa, Asia, Australia, and North and South America (Wang et al. 2017).

This paper concerns one such system, the East Asian summer monsoon (EASM), which covers a domain stretching over the South China Sea, East China, Japan and the Korean peninsula. It is characterised by an abrupt reversal of low-level winds over the South China Sea during May, and the subsequent establishment of a quasi-stationary rain band (the ‘Meiyu’, ‘Baiu’ or ‘Changma’) which propagates northwards in distinct phases through the summer, before retreating southwards in August (Ding and Chan 2005). The EASM is a highly complex system, with many influencing factors e.g. variability in the West Pacific subtropical high; the south Asian High, the subtropical Eurasian jet stream and Pacific and Indian ocean SST anomalies (Ding and Chan 2005).

Millions of people across East Asia are affected by the monsoon: the rainfall it brings accounts for around half of

✉ John W. Rostron
john.rostron@metoffice.gov.uk

¹ Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK

² LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

annual totals over the region, and interannual variations are typically around 30% (Sperber et al. 2013) of the seasonal means, with potentially serious consequences through flooding, drought and impacts on water supplies, agriculture and hydroelectricity generation. Providing predictions of the EASM and its variations is thus vitally important, and much work has been done, particularly on seasonal timescales (Wang et al. 2015). Given the complexity in modelling the system, forecasting has typically been based on statistical relationships. More recently, however, dynamical modelling using the Met Office's GloSea5 system has been shown to be skilful over China, and has formed the basis of a forecast for summer rainfall over the Yangtze River basin (Li et al. 2016; Bett et al. 2018). This region has been a major focus of forecasting and climate prediction research as it is particularly sensitive to interannual variability, with past flooding events impacting the lives of hundreds of millions of people. It is also an important economic region, including some of China's largest cities and being a centre for key industries.

On longer timescales, it is important to understand the role of a changing climate on the EASM. Globally, monsoon activity, variability and the strength of teleconnections to ENSO are expected to increase (Hsu et al. 2012, 2013), whilst Kitoh et al. (2013) found the EASM to respond strongly to warming (compared to other monsoon regions), particularly for metrics of heavy precipitation. Detailed studies looking at moisture budget decompositions over the region have revealed that the precipitation increase is largely driven by moistening of the atmosphere, but that uncertainties in this response are mainly due to uncertainties in circulation: both in the background state and responses to climate change (Zhou et al. 2018; He and Li 2019; Chen et al. 2020b; Zhang et al. 2021).

Given the highly complex nature of monsoon systems, including the EASM, it is crucial to test the robustness of these projected changes to various sources of uncertainty e.g., modelling uncertainty, scenario uncertainty and internal variability (Zhou et al. 2020). Much of the work on projected changes has focussed on multi-model ensembles (e.g., CMIP3, 5 and 6), which bring together the latest model configuration from different centres in 'ensembles of opportunity'. These ensembles sample uncertainties in modelling structures e.g., in their resolution, complexity, parameterisation schemes.

In this paper we will evaluate the present-day performance and twenty-first century changes of the EASM using models that sample a different source of uncertainty: that arising from uncertainties in the values of model parameters. We do this using an alternative approach to ensemble creation, where members share the same model structure but differ in the values they take for uncertain model parameters. Such 'perturbed parameter ensembles' (PPEs) have been used in the study of monsoon systems (Yang et al. 2015;

Huang et al. 2020), as well as in many other research areas e.g. for present-day climate performance (e.g. Yokohata et al. 2013; Sexton et al. 2019, 2021); climate feedbacks and sensitivities (e.g. Sanderson 2011; Collins et al. 2011; Karmalkar et al. 2019; Rostron et al. 2020; Tsushima et al. 2020); emergent constraints (e.g. Wagman and Jackson 2018); and aerosol forcing (Regayre et al. 2018; Johnson et al. 2018).

One key strength of PPEs is their ability to highlight structural behaviours of a model: behaviours (e.g., biases or future changes) that are common to most (if not all) parameter combinations and that cannot be removed through parameter tuning. Conversely, PPEs also identify those aspects of a projections that *are* susceptible to tuning. PPEs are also an excellent tool for parameter sensitivity analyses, where links between model parameter settings and model outputs can be studied to help identify the key processes driving changes in the model. In this paper, we utilise these strengths to add detail to our assessments of the biases and future changes in the EASM.

The PPE studied here is based on recent configuration of the Met Office's global coupled model: HadGEM3-GC3.05 (Yamazaki et al. 2021). It comprises 20 variants of HadGEM3-GC3.05 which were generated through simultaneous perturbations made to 47 model parameters (across 7 atmospheric parameterisation schemes), chosen to sample key parametric uncertainties (see Sect. 2.1). These models formed a key component of the recent UK Climate Projections for land project (UKCP18; Murphy et al. 2018).

To place our PPE analysis in the context of previous studies, we also analyse subsets of models from the CMIP5 and CMIP6 projects. The CMIP ensembles sample different model structures and consequently sample a different source of modelling uncertainty to the PPE. The PPE and CMIP ensembles therefore provide complementary datasets, and considering both types of ensemble allows for a more thorough representation of the uncertainties in model performances and future changes. This is important for testing the robustness of future changes in the EASM, and for providing more comprehensive information to users interested in regional impacts assessments and adaptation work. Indeed, a combination of the PPE and CMIP5 ensembles was used for the global model component of the UK Climate Projections project, UKCP18 (Murphy et al. 2018).

However, given the different nature of the ensembles, comparisons between them can be challenging. Whilst the differences in modelling choices sampled by CMIP5 and CMIP6 make them useful for capturing a wide diversity in model biases and future changes, they cannot reveal the structural behaviours of any individual member. Conversely, whilst the PPE will expose structural biases and future changes, it only does this for a single base model. Consider, for example, the ensemble mean biases. A clear

structural model bias in the PPE will be reflected in its mean, as it would not be removed by parameter perturbations. Each CMIP5 or CMIP6 model could also exhibit clear structural biases (though we wouldn't know without a PPE around each of them), but the effects of these would be suppressed in the ensemble mean due to cancellation across the different model structures (unless it was a bias common across those models e.g., the double ITCZ bias). So, whilst it is advantageous to consider both types of ensembles, we must be cautious when directly comparing them.

Our analysis will focus on performance and future changes for leading-order metrics of low-level winds and precipitation. Whilst these metrics will not represent all the complexities of the EASM, they will encompass many of its key features. For example, climatological low-level winds are crucial for capturing the correct flow of moisture through the region, whilst seasonal cycles of precipitation will be sensitive to the northward propagation of the Meiyu rain band. Numerous previous studies have explored different aspects of the performance of the EASM in HadGEM3 models, from sub-seasonal to climatological timescales (e.g., Li et al. 2016; Rodríguez et al. 2017; Hardiman et al. 2018; Rodríguez and Milton 2019; Martin et al. 2020, 2021). Known relationships between the interannual changes in these variables will also be assessed (Wang et al. 2008). Variability in these relationships on decadal timescales and longer will be of particular interest as they have been noted for their potential use in Met Office seasonal forecasts for the Yangtze river basin (Martin et al. 2020).

The remainder of the paper is ordered as follows: Sect. 2 provides a summary description of the design of the PPE and the sub-selection of CMIP5 and CMIP6 members, along with the methods used to analyse precipitation regionally over China and a metric used for low-level circulation over east Asia. In Sect. 3 we analyse the present-day performance of the PPE, CMIP5 and CMIP6 models, looking at mean state biases and variability of low-level winds and precipitation; seasonal cycles of precipitation; and the relationships in interannual anomalies related to the EASM (including ENSO). Section 4 covers the future changes in these variables, with Sect. 4.2.1 focussing on understanding the drivers of the twenty-first century precipitation response for the Yangtze River basin, including using a detailed moisture budget analysis. In Sect. 5 we explore the sensitivities of EASM circulation metrics to the perturbed parameters using a causal network analysis. We will discuss our findings in Sect. 6, with a focus on how the information

provided by the PPE may be used in practice. An overall summary is given in Sect. 7.

2 Models and methods

2.1 Base model and parameter perturbations

Here we provide a summary description of the HadGEM3-GC3.05 PPE. In this paper we will refer to this as 'the GC3.05-PPE', or simply 'the PPE'. Further details, including a description of the GC3.05 base model, can be found in Yamazaki et al. (2021).

The GC3.05-PPE comprises 20 variants of the UK Hadley Centre Unified Model HadGEM3-GC3.05 model, which is closely related to the GC3.1 configuration submitted to CMIP6 (Williams et al. 2018). Each ensemble member has a horizontal resolution of approximately 60 km at mid-latitudes (called 'N216') and was run for a 200-year period from 1900 to 2100, using CMIP5 historical forcings and future scenarios consistent with RCP8.5 emissions (accounting for carbon cycle uncertainties). Flux adjustments were applied to each member, in order to mitigate the effects of long-term SST (and salinity) biases on the projected regional changes (Murphy et al. 2018; Yamazaki et al. 2021).

Each PPE member is distinguished by taking a unique set of values for 47 model parameters across 7 parameterisation schemes from the atmosphere, land and aerosol model components (A full description of the parameters perturbed in this PPE can be found in Table 1 in Sexton et al. 2021). The initial distributions of parameter values were chosen to target key modelling uncertainties, through an elicitation exercise with model experts. The parameter values ultimately used for the 20 PPE members were selected through a multi-stage filtering process, based on the plausibility of their representation of the climate, and on the diversity of their climate change responses.

The latter was assessed using idealised forcing experiments in atmosphere-only simulations, where diversity in climate feedbacks, aerosol and CO₂ forcings, and regional precipitation and temperature responses were targeted (Sexton et al. 2021). The plausibility of the variants was assessed in a variety of historical and present-day experiments e.g. using large-scale mean climate performance in 5-day and 5-year atmosphere-only experiments, as well as qualitative assessments of circulation, surface air temperature and precipitation over the North Atlantic and UK (Sexton et al. 2021). Further screening was applied to the variants run as fully coupled simulations. An initial

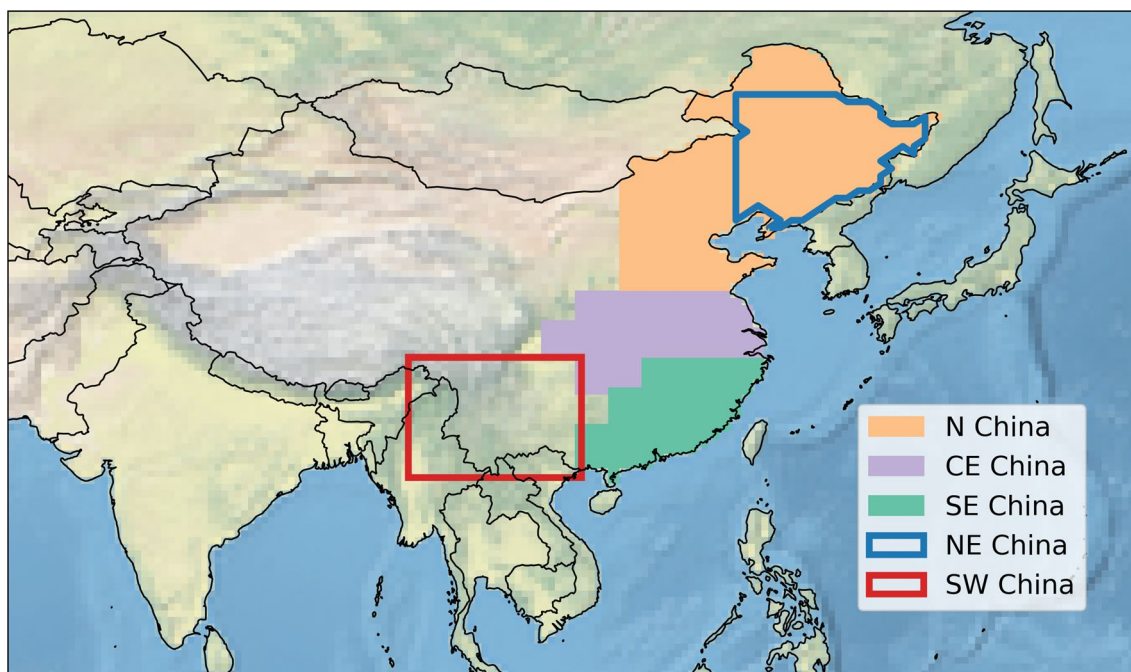


Fig. 1 Regions of China used in our analysis. The shaded regions were selected based on the K-means clustering algorithm described in Sect. 2.3. These are North China (N China; orange); Central-East China (CE China; purple) and Southeast China (SE China; green).

Additional regions used in our analysis, covering Northeast China (NE China; blue box) and Southwest China (SW China; red box), are also shown

ensemble of 25 members was reduced, first to 20 members and then to 15, based on criteria such as: numerical stability; the strength of the Atlantic meridional overturning circulation (AMOC); historical trends in northern hemisphere surface air temperatures; and climatological temperature biases (Yamazaki et al. 2021).

The final 15 members were selected for use in the UKCP18 projections (Murphy et al. 2018), but for this study we are interested in exploring a diverse range of model behaviours, with a focus on a different region (East Asia). So, we choose to use the 20 members selected after the first stage of coupled screening and which were run for the full 200-year period.

One of the 20 PPE members we analyse uses tuned parameter values i.e. the same parameter values used for the HadGEM3-GC3 model (Williams et al. 2018). We refer to this as the ‘standard’ PPE member.

2.2 CMIP5 and CMIP6 models

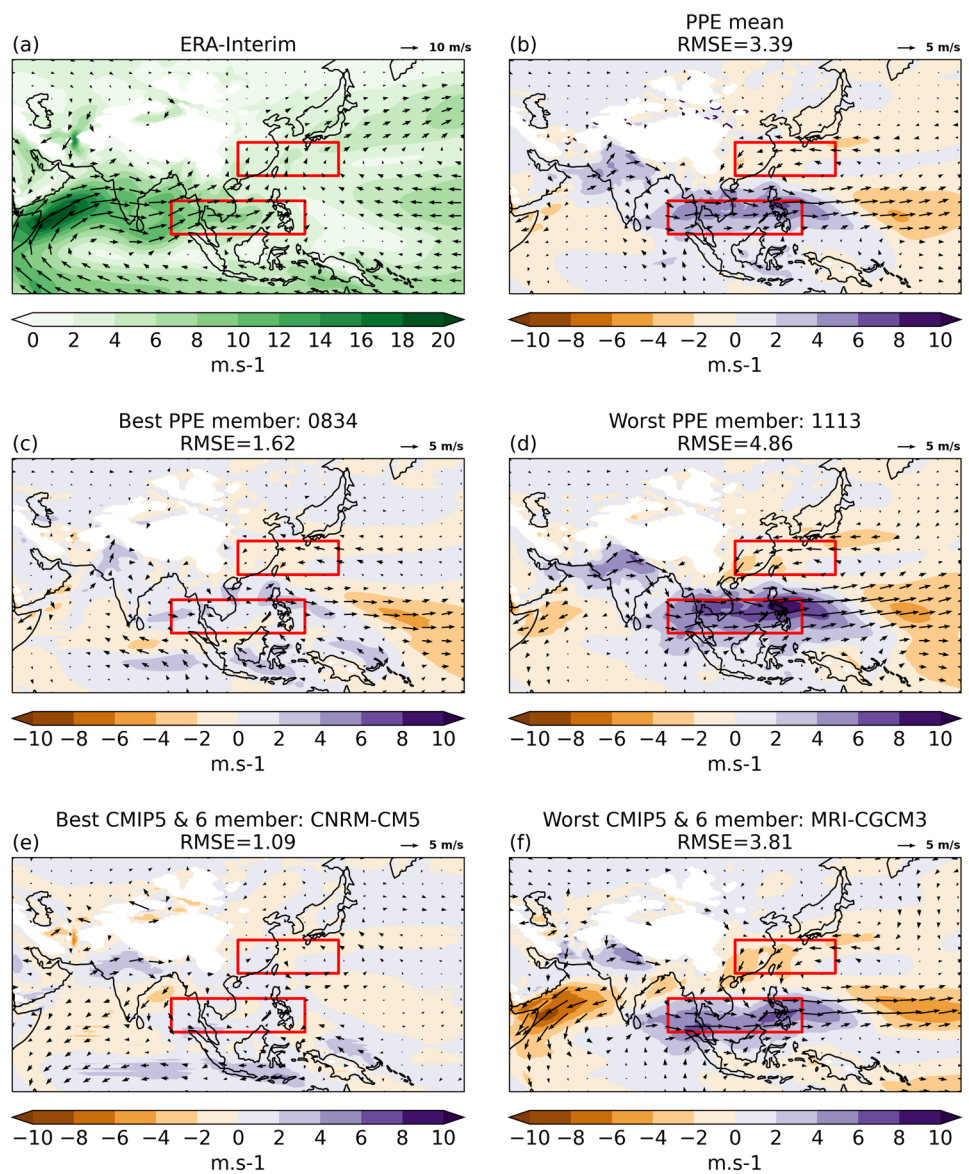
We also assess a subset of CMIP5 and CMIP6 global coupled models, which allow us to sample uncertainties in different model structures, in contrast to the parametric modelling uncertainties sampled by the PPE. (We refer to these subsets as the ‘CMIP5’ and ‘CMIP6’ ensembles

when considered separately, and the ‘CMIP’ ensembles when considered together.) The PPE and CMIP datasets are complementary: CMIP models will provide a useful context in which to place our assessment of the PPE, and consideration of the results from all three ensembles is recommended for users of the projections (but note the caveats regarding direct comparisons discussed in the Introduction).

The CMIP5 subset is formed from members which were selected for the UKCP18 project, based on a qualitative assessment of key aspects of global and European/UK climate, along with a screening of very closely related models (see Murphy et al. 2018 for the details of the selection methodology). We use 12 of these 13 selected models – the EC-EARTH model was not used as not all of the required data was available at the time of our analysis. For the historical period, the CMIP5 forcings are the same as those used for the PPE, whilst for the future period, concentrations from the RCP8.5 pathway are used (note the slight difference to the PPE here, as carbon cycle uncertainties are not sampled for the CMIP5 models).

For our CMIP6 subset we selected the CMIP6 models which were most closely related to our CMIP5 subset, and for which we had data available. In total, our CMIP6 subset contains 10 models. We use CMIP6 historical forcings and the SSP5-8.5 future scenario, as this represents the closest

Fig. 2 JJA mean 850 hPa wind fields. **a** 1980–2015 ERA-Interim climatology, with green shading showing the wind speeds (magnitude of wind vectors). **b–f** Model biases with respect to ERA-Interim for the same period. The arrows represent the zonal and meridional components of the wind biases, while the shading shows the wind speed bias. **b** Mean bias across PPE members. **c** and **d** Biases for the best (0834) and worst (1113) models from the PPE, respectively. **e** and **f** Biases for the best (CNRM-CM5) and worst (MRI-CGCM3) models from the CMIP ensembles, respectively. These were selected using RSME values for the regions used to calculate the RWFI (red boxes)



equivalent to the RCP8.5 scenario used for the PPE and the CMIP5 subset.

All members of the CMIP subsets are re-gridded from their native resolution to that of the PPE members (60 km at mid-latitudes) to facilitate the comparison of these ensembles. A table of the CMIP5 and CMIP6 models used here is given in Appendix A (Table 3).

2.3 Definition of regions

Part of this study into the East Asian summer monsoon will involve an assessment of precipitation across China. As noted in the Introduction, the EASM has a complex spatial

and temporal structure, with the quasi-stationary Meiyu rain band influencing different regions through the season. We therefore separate China into regions, based on areas that display similar characteristics of precipitation variability.

We do this using a K-Means clustering algorithm (Wilks 2011), where climatological monthly anomalies of precipitation are calculated for each grid box over China and boxes with similar annual cycles are grouped together into 3 groups. For this analysis we used GPCP observational data (covering 1980–2014; Adler et al. 2003), and considered land points only. The three regions resulting from

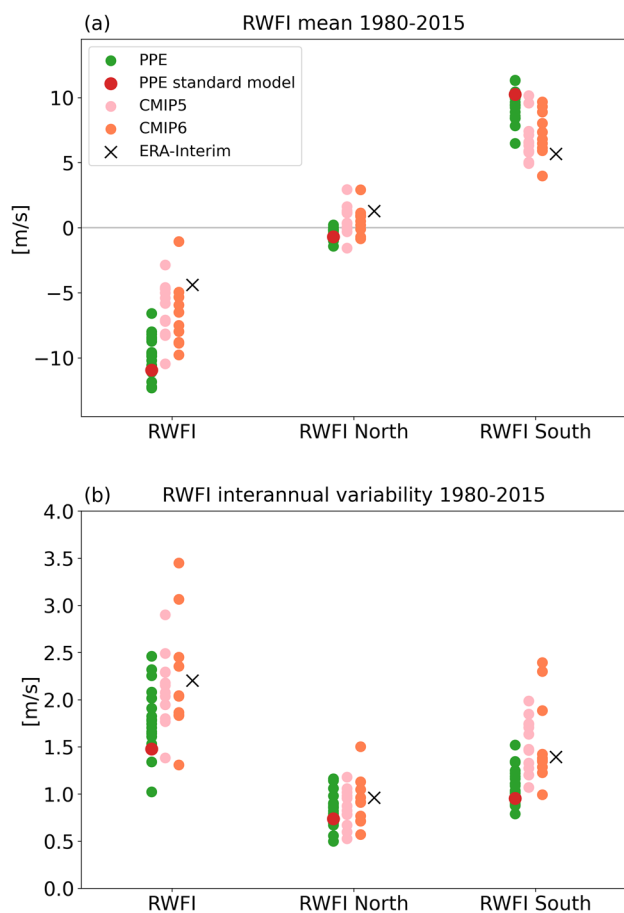


Fig. 3 **a** Mean climatologies (1980–2015) for RWFI and its north and south components (see definition in Sect. 2.4.1). PPE members (green points) are shown alongside CMIP5 (pink) and CMIP6 (orange) models. ERA-Interim values are shown as black crosses. Values for the PPE standard member are shown by the red points. **b** RWFI interannual variability (after removal of a linear trend) over the same period

this clustering were largely continuous, although some grid boxes fell into a different region from their neighbours, particularly near the borders between regions. Because of this, we manually adjusted the regions so they were completely continuous (i.e., no ‘floating’ grid boxes), but still reflected the broad regions selected by the K-Means algorithm. Additionally, one of the regions covered a very large domain, which included all parts of China except the Southeast and Central-East. Much of this domain is not affected by the monsoon, so we limited it to only include points eastward of 100E.

The three regions selected by this analysis are shown in Fig. 1. We label these ‘Southeast’ (SE), ‘Central-East’ (CE) and ‘North’ (N) China, and they align with the north–south propagation of the EASM. We also define 2 further regions:

one for ‘Northeast’ (NE) China, covering the important maize-growing region; and another for ‘Southwest’ (SW) China, where previous versions of Hadley Centre models have consistently shown wet biases in the summer (Rodríguez et al. 2017).

2.4 Metrics

2.4.1 Reversed Wang and Fan Index (RWFI)

Indices are used widely in climate science to quantify features of the climate system in a simple way (e.g., circulation patterns). In this study, we focus on one index used in studies of the East Asian summer monsoon—the ‘reversed Wang and Fan index’ (RWFI). This is defined, using summer (JJA) means, as:

$$RWFI = u_{850}(22.5^{\circ} - 32.5^{\circ}N, 110^{\circ} - 140^{\circ}E) - u_{850}(5^{\circ} - 15^{\circ}N, 90^{\circ} - 130^{\circ}E)$$

(See the red boxes in Fig. 2.) This index reflects the low-level shear vorticity over the region and was initially used to quantify variability in the western North Pacific summer monsoon (Wang and Fan 1999). The complexity of the East Asian summer monsoon precludes any index from capturing all aspects of the system (Wang et al. 2008). However, Wang et al. (2008) found the RWFI correlates very well with the first multivariate EOF for precipitation, surface level pressure and winds over China, and thus provides a simple metric to capture some key features of East Asian summer monsoon variability. We therefore use the RWFI metric as our leading-order metric to study low-level circulation for the EASM.

2.4.2 Nino-3.4 index

In our analysis of the connections between modes of variability affecting the EASM (Sect. 3.3) we consider the role of ENSO and its impact on the EASM circulation. We quantify ENSO using the Niño-3.4 index, which is defined using the long-term anomalies of the monthly-mean SST averaged over the Niño-3.4 region ($5^{\circ}S - 5^{\circ}N, 170^{\circ}W - 120^{\circ}W$), where we use a baseline climatology period from 1950 to 2006. We then smooth the time series using a 5-month window and normalise using the standard deviation of the smoothed time series over the climatological period.¹

¹ The definition of this Niño-3.4 index can be found in the ‘Technical Notes’ section of <https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni>.

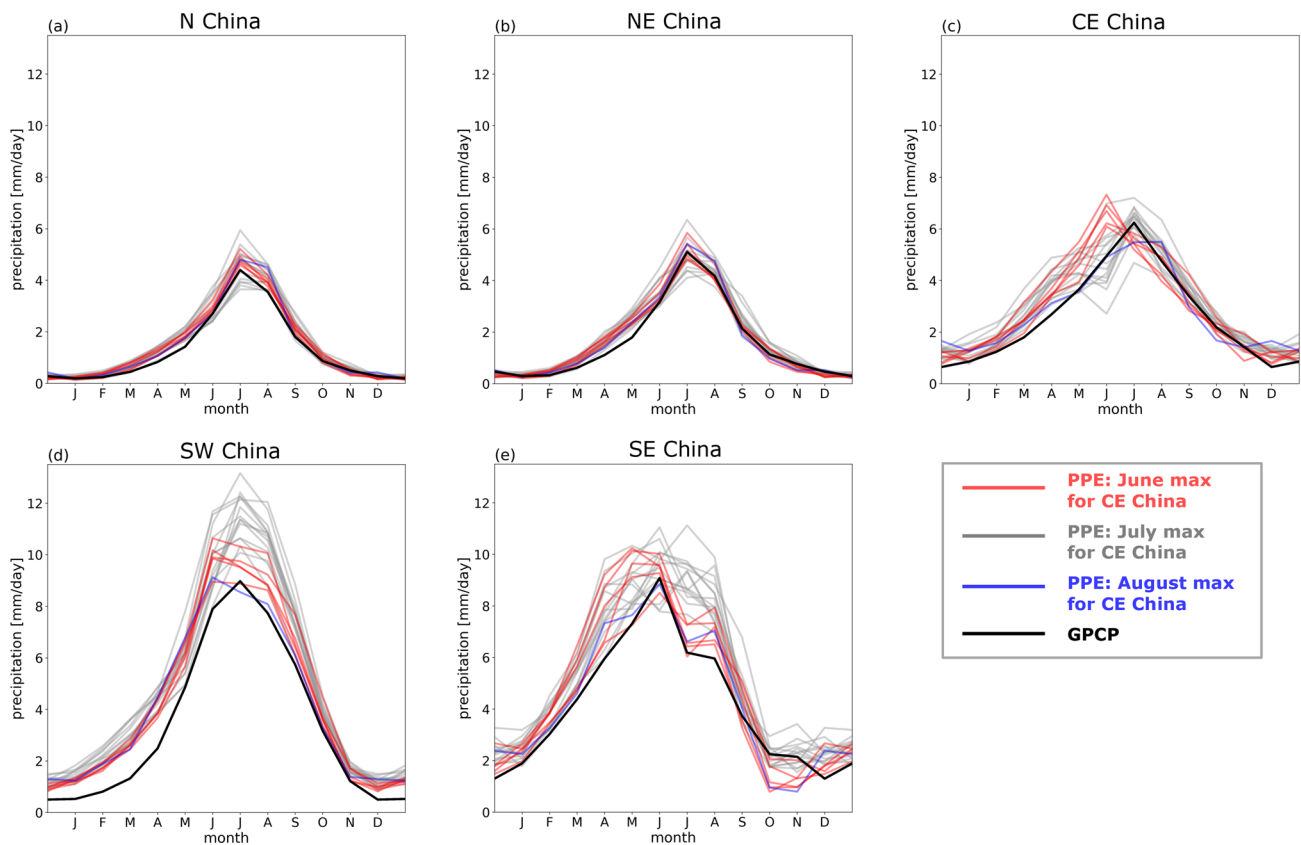


Fig. 4 Annual cycles of precipitation for regions in China (see Sect. 2.3 for definitions). Monthly climatologies (based on 1980–2014 means) are shown for observations from GPCP (black lines) and for each PPE member. The colours used for each PPE member indi-

cate the month when CE China precipitation reaches its maximum: red indicates a maximum in June; grey, a maximum in July and blue, a maximum in August

3 Present-day performance of the EASM

3.1 850 hPa circulation

We start with an assessment of the lower-level (850 hPa) circulation over the region. Figure 2a shows the JJA mean climatological 850 hPa circulation from reanalysis (ERA-Interim; Dee et al. 2011) for the period 1980–2015. It is characterised by an airflow along the Somali coast, across the Arabian Sea (a moisture source for the Indian monsoon) and the Bay of Bengal and into the South China Sea. This circulation develops rapidly during May (Ding and Chan 2005) and persists through NH summer. The Bay of Bengal, the Philippine Sea and the South China Sea are key moisture sources for the EASM (Zhou 2005).

The PPE mean bias field (Fig. 2b) shows substantial errors in this circulation, with a cyclonic bias over Southeast Asia and the west Pacific. The westerly bias centred on the South China Sea, resulting from an over-extension of the Indian monsoon flow, is particularly strong. This bias is

characteristic of a known circulation error that has affected previous generations of Hadley Centre models (Ringer et al. 2006; Bush et al. 2015) and is present in atmosphere-only (as well as coupled) simulations, and across resolutions (Rodríguez et al. 2017; Chen et al. 2018). It is associated with biases in the representation of the western North Pacific subtropical high (WNPSH), which tends to be too weak and shifted too far to the east in these models (Chen et al. 2018).

Figure 2c–f show examples of the ‘best’ and ‘worst’ members from the PPE and the CMIP ensembles, as measured by root-mean-squared errors of the zonal wind component over the area highlighted by the red boxes (see Sect. 2.4.1). The cyclonic bias pattern in the PPE mean is also seen in the best and worst PPE members (labelled ‘0834’ and ‘1113’, respectively; Fig. 2c and d). In the latter case, the errors over the South China Sea reach up to 9.2 ms^{-1} but the errors are clearly much reduced for member 0834. A similar bias is seen for the worst amongst the CMIP models (MRI-CGCM3, from CMIP5; Fig. 2f), but the best model here (CNRM-CM5, also from CMIP5) performs well over the whole domain (Fig. 2e).

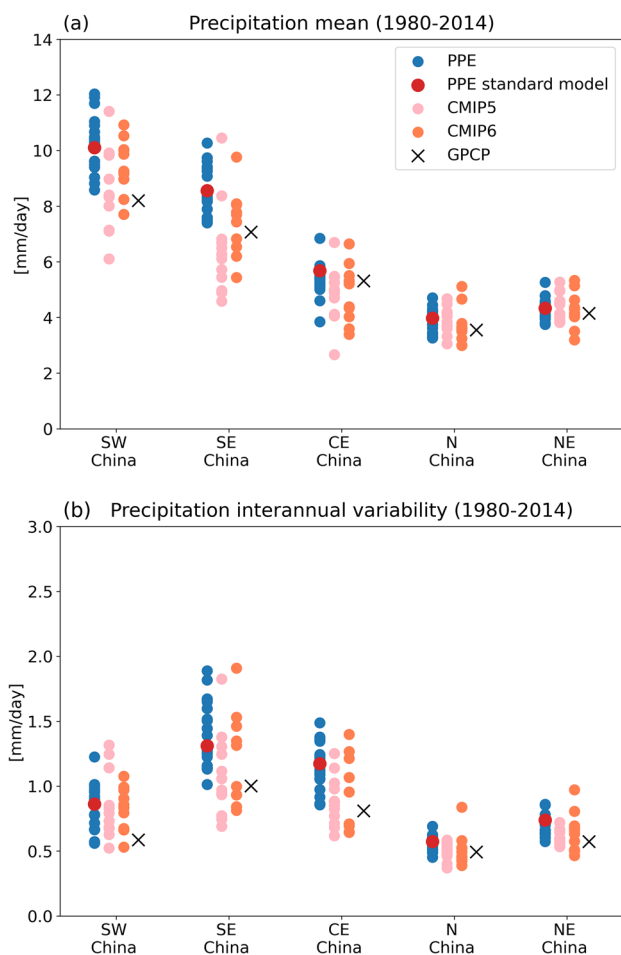


Fig. 5 **a** JJA mean precipitation climatologies (1980–2014) for regions in China (see Sect. 2.3 for definitions). PPE members (blue points) are shown alongside CMIP5 models (pink points), CMIP6 models (orange points) and observations from GPCP (black crosses). Values for the PPE standard member are shown by the red points. **b** Interannual variability of JJA precipitation (after removal of a linear trend) over the same period

These features are seen across the PPE and CMIP ensembles, as shown in the 1980–2015 mean climatologies of the RWFI (Fig. 3a), which measures the 850 hPa vorticity over the region (as sampled by the red boxes in Fig. 2; see Sect. 2.4.1). The RWFI climatologies for the PPE members show a structural negative bias for the PPE, with the cyclonic nature highlighted by consistently negative biases for RWFI-North and positive biases for RWFI-South.

Negative RWFI biases are also seen in most of the CMIP models (Fig. 3a). This is consistent with previous studies of the EASM circulation in CMIP5 models, where the low-level westerly jet is typically found to be too strong, and is associated with northeastward shifts in the WNPSH (Song and Zhou 2014a, b). However, there are instances of positive RWFI biases e.g., ACCESS1-3 (CMIP5) and CNRM-CM6-1

(CMIP6). In the PPE and the CMIP ensembles, the spread in RWFI values results mainly from the southern component: understanding the processes and structural changes driving the spread in this region could be crucial for resolving this model error (Bush et al. 2015; Martin et al. 2021).

Figure 3b shows the interannual variability in RWFI (and its components) for each ensemble member. All 3 ensembles span a range of variability, from 1.0 to 3.5 ms^{-1} , which includes the value from the ERA-Interim reanalysis. PPE members typically show less variability than the reanalysis in all components. Figure 3b also shows that CMIP models are typically less variable than the reanalysis for RWFI-North, but there are examples of enhanced variability for RWFI and RWFI-South. One notable model is CESM2 (from CMIP6) which has the most overestimated variability in all 3 RWFI components but is amongst the best of all the models for the mean climatologies for these components. This highlights the importance of considering multiple metrics of when judging the performance of models.

3.2 Precipitation

As noted in the Introduction, a key characteristic of the EASM is the northward progression of the Meiyu rain band across the region, from June to August. Accordingly, we start by evaluating the performance of the PPE for precipitation using the annual cycles across the distinct regions defined in Sect. 2.3, to capture the large-scale spatial and temporal characteristics of the EASM precipitation (see Fig. 4).

Across all of the regions, the observed annual cycles (GPCP; Adler et al. 2018) show a continuous increase in precipitation from winter to a peak in summer and a subsequent decrease after the retreat of the EASM. These variations are broadly captured by the PPE, although values for MAM are consistently overestimated by all members, across the regions. This bias was also noted by Rodríguez et al. (2017) for a previous, atmosphere-only configuration of HadGEM3 (GA6; Walters et al. 2017), and is associated with errors in the moisture convergence.

The timings of the peak of the precipitation in summer is well captured across PPE members for N, NE and SW China, but there is a notable spread in CE and SE China. Members that peak prematurely in CE China (i.e., in June rather than the observed peak in July) also tend to have a premature peak in SE China (see the red curves in Fig. 4e), indicating coherence in these timing errors across the PPE. However, this subset tends to capture the JJA mean precipitation in SW China more accurately than the members with the correct peak timing (in July) in CE China. This can be seen by comparing the red curves to the grey curves in Fig. 4c and d.

We show 1980–2014 mean JJA precipitation values for PPE members in Fig. 5a, along with the CMIP ensembles and GPCP observations. Clear wet biases are seen in SW

and SE China, which is consistent with previous configurations of HadGEM3 (Rodríguez et al. 2017). The fact that all PPE members show wet biases in these regions indicates that these biases are structural. No such structural biases are found for N China; NE China, which is an important crop growing region; and CE China, which covers the Yangtze River valley.

Biases in the CMIP models also vary spatially. Despite the differences in their constructions (as noted in the Introduction), the PPE, CMIP5 and CMIP6 ensembles all span the observed climatology in CE, N and NE China. In SW and SE China there are examples of positive and negative biases in both CMIP5 and CMIP6. Notably, in SW China all but one of the CMIP6 models has a wet bias (similar to the PPE), but in SE China all but two of the CMIP5 models has a dry bias (in contrast to the PPE). In these regions the value of combining these ensembles to more thoroughly capture a diversity in performance is clear.

The interannual variability in JJA precipitation is shown in Fig. 5b. The variability for PPE members structurally high across all regions, with values ranging between 0.92 and 2.09 times the observed standard deviation values. CMIP models also tend to overestimate the variability, but there are examples of models with too little variability, reflecting a more diverse sampling of precipitation variability in the CMIP models compared to the PPE.

We have also calculated precipitation variability scaled by the climatological mean for each region (not show) and found that these were also consistently overestimated in the PPE. The reasons why the PPE overestimates precipitation variability is not clear, but some of this may be driven through relationships with the monsoon circulation (Wang et al. 2008; also see Sect. 3.3). An overestimation of precipitation variability could be driven by too much variability in the circulation acting through this teleconnection, or from errors in the teleconnection itself. However, we do not find evidence of this in the PPE. As shown in Fig. 3b, the variability of the RWFI in the PPE is reasonable (even slightly underestimated). We have also analysed the relationships between circulation and CE China precipitation in the PPE (see Sect. 3.3) but find that values of the fraction of variance of precipitation explained by RWFI are between 0.02 and 0.35 (the observed value is 0.29), indicating that other influences are driving our overestimated precipitation variability in this region.

3.3 Variability relationships

Well known relationships exist between interannual anomalies in EASM circulations, summer precipitation over East China, and ENSO (e.g., Ronghui and Yifang 1989; Wang et al. 2000, 2008; Chang et al. 2000; Wu et al. 2003; Zhou 2005; Li et al. 2016). Here, we explore how these

Table 1 Relationships between CE China precipitation (prC), EASM circulation (using the RWFI) and ENSO (using the Niño-3.4 index) for observations (top row) and the PPE (remaining rows)

Member		prC-RWFI	RWFI-Niño-3.4
X	OBS	0.200 [0.110]	0.970 [0.525]
A	0000	0.392 [0.262]	0.371 [0.409] ^x
B	0090	0.349 [0.232]	0.030 [0.416] ^{x†}
C	0605	0.311 [0.255]	0.262 [0.520] ^x
D	0834	0.117 [0.309] ^x	− 0.193 [0.371] ^{x†}
E	1113	0.253 [0.141]	0.956 [0.633]
F	1554	0.246 [0.230]	0.538 [0.382]
G	1649	0.281 [0.170]	− 0.322 [0.560] ^{x†}
H	1843	0.305 [0.234]	0.169 [0.488] ^{x†}
I	1935	0.344 [0.215]	0.205 [0.520] ^{x†}
J	2089	0.253 [0.164]	1.237 [0.467] ^x
K	2123	0.224 [0.246] ^x	− 0.051 [0.525] ^{x†}
L	2242	0.257 [0.240]	− 0.231 [0.614] ^{x†}
M	2305	0.236 [0.200]	0.109 [0.557] ^{x†}
N	2335	0.146 [0.212] ^x	0.339 [0.605] ^x
O	2491	0.330 [0.212]	0.730 [0.486]
P	2753	0.562 [0.270] [†]	0.073 [0.420] ^{x†}
Q	2832	0.090 [0.223] ^x	0.498 [0.431]
R	2868	0.260 [0.166]	0.474 [0.742] ^x
S	2884	0.154 [0.143]	0.423 [0.611] ^x
T	2914	0.254 [0.165]	0.400 [0.504] ^x

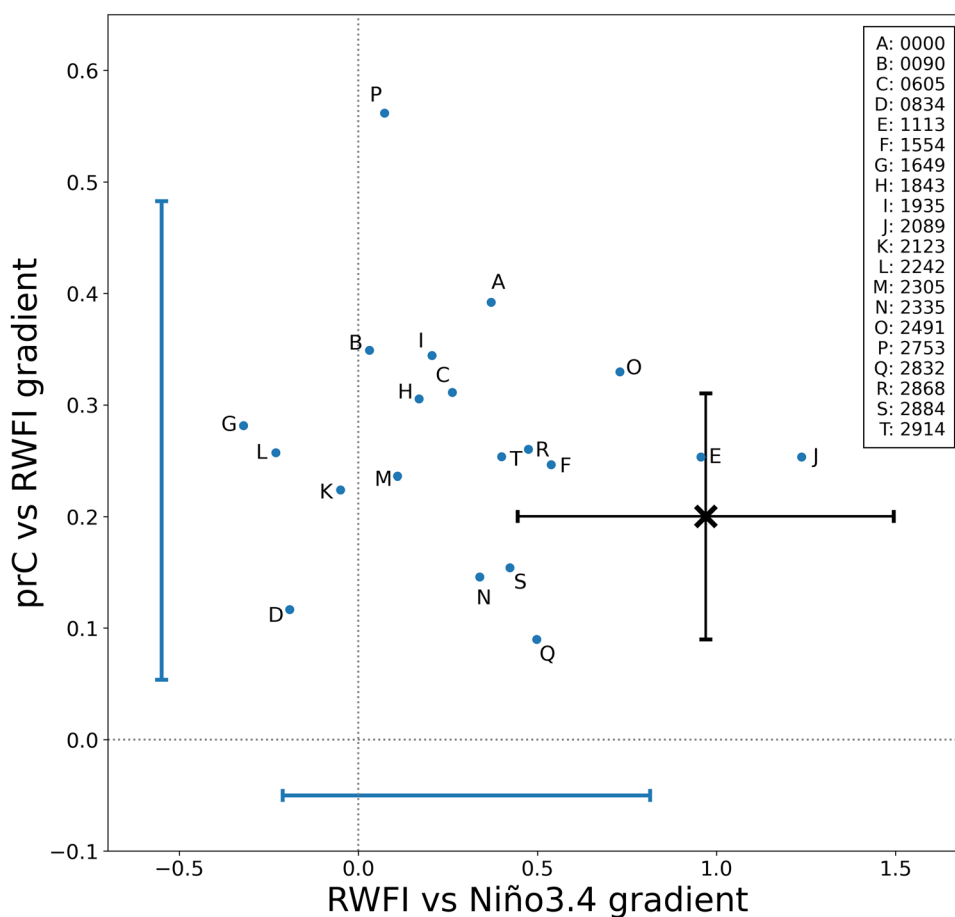
Values are calculated as the slope of the linear relationship between interannual anomalies of these metrics for the period 1980–2014. For prC and RWFI we use JJA means, whilst for the Niño-3.4 index we use DJF means from the preceding winter. Values in brackets give the uncertainties capturing the 95% confidence range. Crosses (^x) denote slopes which are not significant, whilst daggers ([†]) denote a significant difference from the observed value. Values with neither a cross nor a dagger therefore denote significant slopes which are consistent with the observed value. (Significance is assessed at the 5% level)

relationships are represented in the PPE by focusing on the connection between the RWFI, JJA mean precipitation for CE China and ENSO.

We focus on CE China, as this coincides with the Yangtze River valley—a region where the impacts of interannual rainfall changes can be great, but also where the strength of these relationships is strong, opening up opportunities for exploiting predictability in models (Bett et al. 2018; Martin et al. 2020). Also, as described in Sect. 3.2, our PPE validates reasonably well against observations in this region.

We start with the relationship between JJA mean precipitation for CE China ('prC') and the RWFI. Wang et al. (2008) showed a strong relationship between the RWFI and the first multivariate EOF of EASM variability, where positive RWFI anomalies are associated with enhanced precipitation over a region which coincides with our CE China region (see Fig. 2a in Wang et al. 2008). These anomalies are characteristic of an enhanced south-westerly flow over

Fig. 6 Relationships between interannual anomalies in JJA mean precipitation (for CE China), RWFI and Niño-3.4, based on the period 1980–2014. Values for the gradient of the relationship between CE China precipitation and RWFI are plotted on the y-axis, whilst RWFI–Niño-3.4 gradient values are plotted on the x-axis. The black cross is for the observed/reanalysis values GPCP and Era-Interim, while the PPE members are shown as blue points (with labels for each PPE member shown in the legend). The black error bars show the uncertainties capturing the 95% confidence range on these gradients for observations/reanalysis. The blue error bars show the equivalent for the mean uncertainty across the PPE members



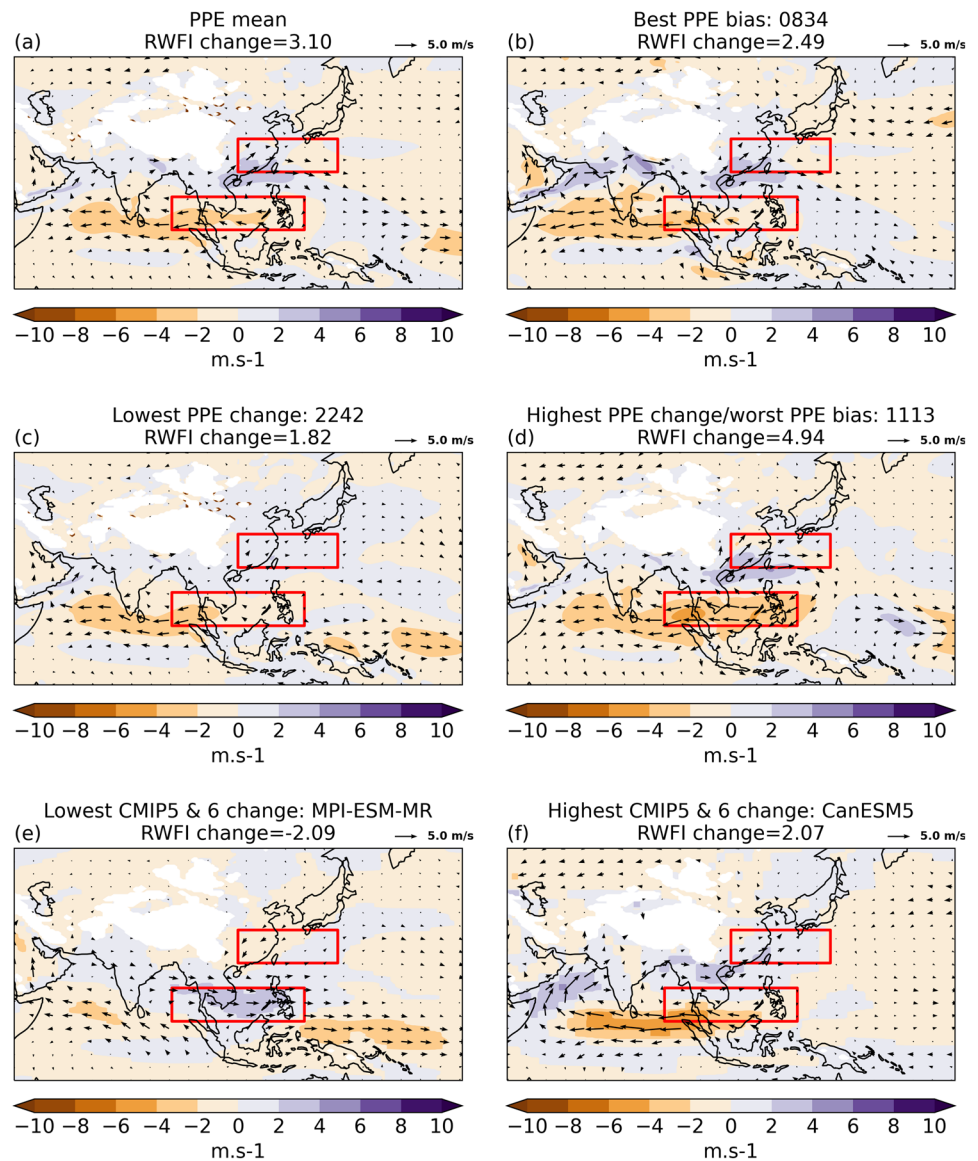
the South China Sea (and a reduction in the zonal wind as measured by RWFI-S), associated with a south-westward extension of the WNPSH, a weakened monsoon trough in the western North Pacific and a southward shift of the upper East Asian jet stream (Chang et al. 2000; Zhou 2005; Wang et al. 2008).

A simple way to characterise this relationship is to evaluate the slope of a simple linear fit to interannual anomalies of CE China precipitation against the RWFI index. We show these values in Table 1 (column ‘prC–RWFI’). The ‘observed’ (‘OBS’) value is derived from 35 years of data from GPCP (for prC) and ERA-Interim (for RWFI) where these two datasets overlap (1980–2014). The observed value of 0.200 ± 0.110 confirms the expected relationship between these quantities is significant. (Uncertainties given are for the 95% confidence range and significance is tested at the 5% level.) The remaining rows in Table 1 give the equivalent values for the PPE members. Most PPE members also exhibit significant relationships, with slopes that are indistinguishable from the observed relationship (at the 5% level). This can also be seen in the y-values in Fig. 6. The exceptions to this are members 0834, 2123, 2335 and 2832, for which our evidence isn’t strong enough to distinguish their slopes from zero; and member 2753, which has a steep slope

that is not consistent with the observations. Most notable amongst these is member 0834 (see point labelled ‘D’ in Fig. 6). This member has the smallest circulation biases in the PPE (see Sect. 3.1), but also has one of the least sensitive and least realistic prC–RWFI relationships.

Interannual variability in the EASM is known to be strongly influenced by the El Niño–Southern Oscillation (ENSO) and many studies have explored the potential mechanisms linking ENSO to anomalies in EASM circulation. Wang et al. (2008) showed that the peak of a lead-lag correlation between their first multivariate EOF, which exhibits an anomalous subtropical high in the west Pacific, and the Niño-3.4 index, occurs in the preceding winter. That is, the anomalous anticyclone, which is associated with an enhanced south-westerly flow over the South China Sea (decreased RWFI-S and increased RWFI) occurs in the summer after an El Niño. This has been linked to positive feedback mechanisms in the west Pacific and the Indian ocean, where ENSO-induced SST anomalies drive Rossby (west Pacific) and Kelvin (Indian ocean) waves, which reinforce the anticyclonic circulation, allowing it to persist into the summer (Wang et al. 2000; Xie et al. 2016; Xie and Zhou 2017; Hardiman et al. 2018).

Fig. 7 Changes in JJA mean 850 hPa winds for 2070–2099 vs 1980–2009. The arrows show the changes in the zonal and meridional components of the wind, while the shading shows the change in the wind speed. **a** PPE mean. **b** PPE member with best circulation performance (0834). **c** and **d** PPE members with the lowest (2242) and highest (1113) changes in RWFI, respectively. **e** and **f** CMIP members with the lowest (MPI-ESM-MR) and highest (CanESM5) changes in RWFI, respectively



We characterise this relationship in a similar way to the prC-RWFI relationship – using the slope of a simple linear regression between the RWFI and the Niño-3.4 index (see definition in Sect. 2.4.2). We use DJF averages for the Niño-3.4 index and regress against the RWFI for the following summer, to capture the peak of the lead-lag correlation described in Wang et al. (2008). As with the prC-RWFI relationship, we use the period 1980–2014 to evaluate the slopes (with the data for the Niño-3.4 index starting from December 1979). The results are shown in the third column of Table 1 (‘RWFI- Niño-3.4’). The observed relationship is significant, as expected, with a value of 0.970 ± 0.525 . However, the PPE exhibits different behaviours: for most of the members (15 out of 20) the RWFI vs Niño-3.4 slope is not distinguishable from zero, and 9 of these have distinctly different relationships to the observations. This can be seen

quite clearly in Fig. 6 (x-values), where PPE members generally have smaller RWFI-Niño-3.4 slope values than the observations, with some values even being negative. The exceptions to this are PPE members 1113, 1554, 2089, 2491 and 2832, which are indistinguishable from the observed relationship (at a 5% confidence level). Except for 2832, these members all matched the observed prC-RWFI relationship too, as highlighted by the clustering of these members around the observations in Fig. 6.

We again note the case of member 0834, which has the smallest circulation errors but captures neither the prC-RWFI relationship nor the RWFI-Niño-3.4 relationship (see point ‘D’ compared to the observations in Fig. 6). This highlights the fact that a model with good performance in some respects (e.g., circulation) does not imply it is a better model in general, and that multiple metrics of performance

should be considered when using these models (e.g., for future projections).

4 Future changes in the EASM

Each PPE member was simulated out to 2100 under the CMIP5 RCP8.5 scenario (as outlined in Sect. 2). Here we assess how the EASM responds to this future scenario across our PPE members, in terms of the mean-state and variability of the low-level circulation (Sect. 4.1) and precipitation (Sect. 4.2). In Sect. 4.2.1 we use a more detailed moisture budget analysis for changes in precipitation of the CE China region, to highlight the relative impact of thermodynamic and dynamic changes on the precipitation response.

4.1 850 hPa circulation

Twenty-first century circulation changes, based on differences in 30-year averages around 1995 (1980–2009) and 2085 (2070–2099), are shown in Fig. 7. The PPE mean (Fig. 7a) shows a clear anti-cyclonic change over the region, with a weakened westerly flow over the SE Asian peninsula and South China Sea, and increased south-westerly flow over East China, suggesting an intensification of the EASM circulation. This change is seen consistently, but with varying magnitudes, across individual PPE members (Fig. 7b–d). The mean state changes in the RWFI (and its components) shown in Fig. 8a are consistent with this: westerly (easterly) changes are seen in the north (south) components, and the RWFI change is consistently positive as a result. This systematic change across PPE members suggests it is driven by a structural response of HadGEM3-GC3.05, which the parameter perturbations do not alter significantly.

Amongst the CMIP models there are examples of both anti-cyclonic and cyclonic changes in the region (2 examples are shown in Fig. 7e and f). This is particularly the case for CMIP5; amongst our CMIP6 models all but two have a positive RWFI change, as in the PPE. But without PPEs based on the CMIP models it is not possible to assess whether any of these are systematic responses (as we find for HadGEM3-GC3.05). Interestingly, Chen et al. (2020a) found that constraining CMIP5 models, based on present-day SST patterns associated with uncertainties in projections of the WNPSH, favoured models with a future strengthening of the WNPSH. Such a change is consistent with the circulation changes we have found in the PPE. Even so, Fig. 8a shows that magnitudes of the changes for the CMIP5 (and CMIP6) models are smaller than in the PPE; that is, the PPE appears to have particularly strong circulation changes.

In Fig. 9 we show these circulation changes against the present-day mean values. This shows a cluster of 6 PPE members (with present-day means $> -9 \text{ ms}^{-1}$) which have

comparable performance to the CMIP models. This cluster of PPE members samples future changes at the lower end of the PPE, but that are systematically higher than the CMIP model changes. Again, this highlights the benefit of considering information from across these ensembles, with CMIP5 and/or CMIP6 providing a wide diversity of future behaviours, and the PPE providing more examples of large, but still plausible, future changes. Note though, that the combination of the PPE and CMIP ensembles should not be considered as one entity, so the combined scatter should not be considered as evidence of an emergent relationship without more information e.g., PPEs based on each CMIP member. Figure 9 also shows that, whilst there is little evidence of a relationship between present-day biases and future changes in the CMIP models, there is a significant relationship for the PPE (at the 5% level; $R^2=0.38$), where models with larger present-day biases tend to have stronger future changes.

Figure 8b shows the change in variability (the standard deviation) in the RWFI components. No consistent change in the variability is seen for any component—the PPE, CMIP5 and CMIP6 ensembles have members with increases and decreases in variability, and the magnitude of these changes are similar in each ensemble.

4.2 Precipitation

Future changes in JJA precipitation, for 2070–2099 vs 1980–2009, are shown for our 5 regions in Fig. 10. Positive changes in both the mean state and interannual variability are widespread: all PPE members get wetter in all regions, as do most CMIP5 and CMIP6 members (Fig. 10a). A small number of CMIP models show a slight drying in some regions e.g., IPSL-CM5A-MR in SW, SE and CE China. For some members/regions, the changes are weak and not statistically significant (at the 5% level)—for example CNRM-CM5 does not show a significant change over SW, SE and CE China, whilst a weak drying in CNRM-CM6-1 for SW and CE China is also not significant. Typically, however, the future changes are significant, particularly in the PPE. Variability also typically increases in these regions (Fig. 10c; Zhang et al. 2021), except for a handful of members from each ensemble.

As described in Sect. 3.2, biases vary across the different models and our chosen regions. For climate service applications, users may want to apply bias corrections by analysing the percentage change in summer precipitation to the modelled climatology, which we show in Fig. 10b. In both the percentage and absolute changes, the precipitation changes in the PPE are typically larger than those in the CMIP5 and CMIP6 ensembles: PPE values range from 0.3 to 4.5 mm/day (7 to 57%), whilst CMIP5 values range from -0.7 to 2.9 mm/day (-15 to 36%) and CMIP6 from -0.4 to 3.3 mm/day (-5 to 40%). This is most notable

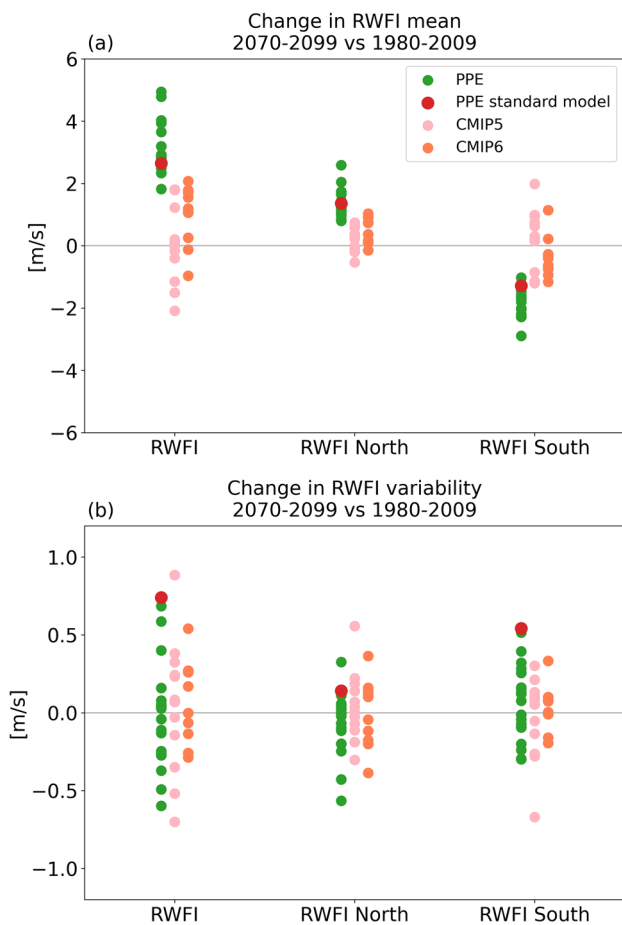


Fig. 8 **a** Change in the mean for RWFI and its north and south components, for 2070–2099 vs 1980–2009. PPE members are shown in green, while CMIP5 models are shown in pink and CMIP6 models are shown in orange. The PPE standard member is shown in red. **b** Equivalent to **a**, but for the change in the interannual variability of RWFI and its components (after the removal of linear trends for the two periods)

in SE and CE China, although there is substantial overlap between the ensembles across the regions.

We emphasise that the structural precipitation responses seen in the GC3.05-PPE may also be present in CMIP models. But without PPEs based around these models we cannot assess this. As was the case for the precipitation biases (Sect. 3.2), considering a combination of these ensembles will clearly be of benefit to users interested in capturing an appropriate level of plausible diversity in precipitation changes over China.

4.2.1 Drivers of precipitation changes

In the previous section we showed that PPE members consistently project increases in precipitation for China over the twenty-first century, but that there is a sizable amount of spread in the magnitude of this change. We now look

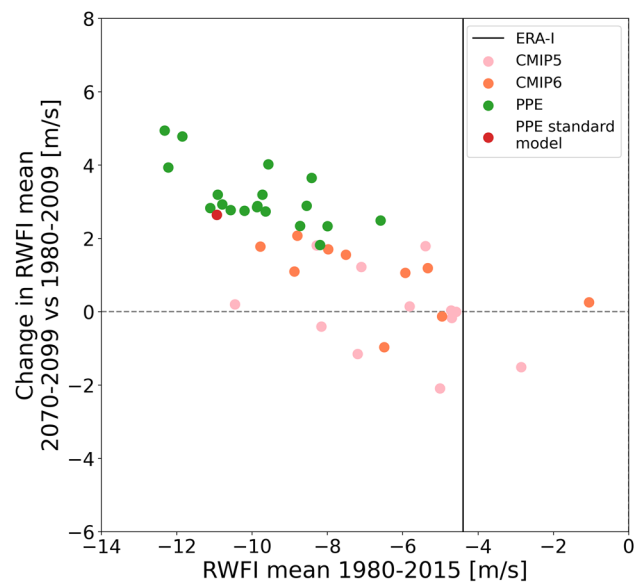


Fig. 9 The change in mean RWFI (for 2070–2099 vs 1980–2009) vs present-day mean RWFI values (1980–2015) for the PPE (green points), CMIP5 models (pink points) and CMIP6 models (orange points). The PPE standard member is shown in red. The present-day value from the ERA-Interim reanalysis is shown with a vertical black line

in more detail at what is driving these changes for the CE China region, starting with a simple assessment of future prC-RWFI relationships, followed by a closer look at changes in the moisture budget. We use this region because, as noted in the Introduction, it has been particularly sensitive to the impacts of climate variability. Additionally, in this paper we have shown that the GC3.05-PPE captures the observed summer precipitation well over the region.

In Sect. 3.3 we showed positive relationships between anomalies in summer precipitation for CE China (prC) and RWFI for the PPE, as well as observations. We have also seen increases in the mean values for both prC and RWFI, so a simple question to ask is: can the PPE’s future changes in prC be explained by the future changes in RWFI applied to the present-day prC-RWFI relationship? This is analogous to the rainfall changes being driven by changes in the large-scale monsoon circulation (to the extent that this is described solely by RWFI), but assuming that any adjustments in precipitable water or the relationship between RWFI and prC are small. The second of these assumptions (stationarity of prC-RWFI relationship throughout the twenty-first century) is also related to changes in the large-scale circulations: for example, future changes in the strength or position of the WNPSH could alter the relation between RWFI and prC.

We tested the stationarity of the prC-RWFI relationship by calculating the gradients for the prC-RWFI linear fits in four 50-year periods across the twenty-first century

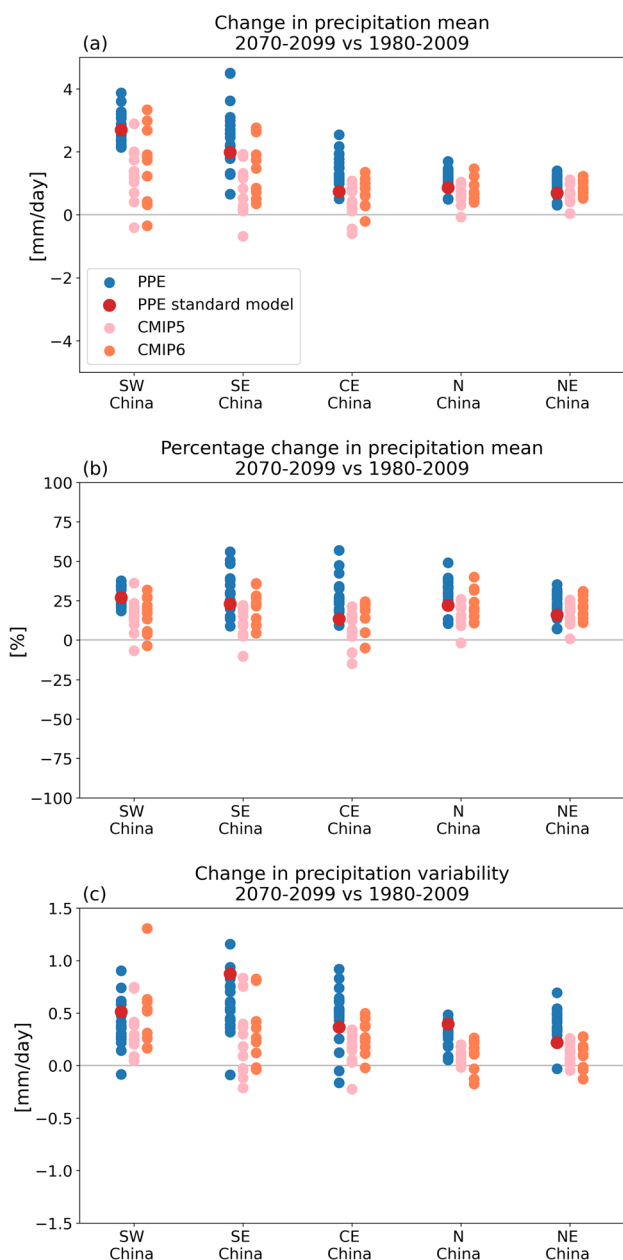


Fig. 10 **a** Change in JJA mean precipitation for regions in China (for 2070–2099 vs 1980–2009) for the PPE (blue points), CMIP5 models (pink points) and CMIP6 models (orange points). The PPE standard member is shown in red. **b** Equivalent to **a**, but with the change expressed as a percentage change from the reference period (1980–2009). **c** Change in the interannual variability of precipitation (after the removal of linear trends for the two periods)

(1989–2039, 2009–2059, 2029–2079 and 2049–2099) and comparing to gradients for 1969–2019. The results are shown in Fig. 11. This shows the prC-RWFI relationships are not constant, with gradient values at the end of the twenty-first century showing little correlation with 1969–2019 values. For instance, member 2242 has a prC-RWFI gradient of 0.21 for 1969–2019 (closely matching the observed value

of 0.20 for 1980–2014), but has a quite different relationship for 2049–2099, with a slope of -0.22 . Conversely, member 2832 has a weak gradient of 0.04 for 1969–2019, but stronger relationships of 0.22 and 0.26 for 2029–2079 and 2049–2099, respectively. The variability in the prC-RWFI relationship suggests that this simple framework is not sufficient for understanding prC changes, and a more detailed analysis of changes in the moisture budget is required.

To do this we analysed daily means of precipitation, evaporation and variables contributing to the moisture convergence—including its decomposition into thermodynamic and dynamic contributions. From these, we determine changes in the components of the moisture budget for CE China, averaged over summer (JJA) for two 30-year periods: 1980–2009 and 2070–2099. The details of these calculations are given in Appendix B. The results of this decomposition are shown in Fig. 12.

The precipitation changes (DP) shown in Fig. 12 are equivalent to those for CE China in Fig. 10. Changes in evaporation (DE) are small, with values ranging from -0.16 to 0.39 mm/day (not shown). Consequently, the precipitation changes are largely driven by changes in the moisture convergence ($R^2=0.84$; see Eq. 9 in Appendix B).

The residual term (DRES) from the decomposition is also shown in Fig. 12. Whilst values are small compared to DP and DconvQ, they are typically negative across the PPE, ranging from -0.52 to 0.06 mm/day. The main driver of this residual is not clear, but there will be contributions from the change in the surface term (see Eq. 1 in Appendix B), from errors introduced through the use of daily means and from errors in the divergence and integration calculations. The size of these residuals places limits on how confidently we can discuss terms in the moisture budget analysis.

Results from the further decomposition of DconvQ into thermodynamic (DTH) and dynamic terms (DDYN; see Eqs. 5–12 in Appendix B) are also shown in Fig. 12. We find that DTH is always positive, as expected from the moistening of the atmosphere in a warmer climate. DDYN also typically contributes positively to DconvQ: only three members have negative dynamic changes, and their magnitudes are small.

In Fig. 13a and b we show correlations between DP-DE and these components. These show that the spread in DP-DE is driven mainly by the dynamic changes, which explain 60% of the variance across PPE members. The importance of changes in circulation here is consistent with results from CMIP5 and CMIP6 (He and Zhou 2015; Zhou et al. 2018; Chen et al. 2020b). In contrast, the thermodynamic changes show little correlation with DP-DE ($R^2=0.01$), and typical values for DRES (indicated by the error bars) are comparatively small.

The dynamic component of this decomposition (DDYN) is a sum of contributions from changes in the mean

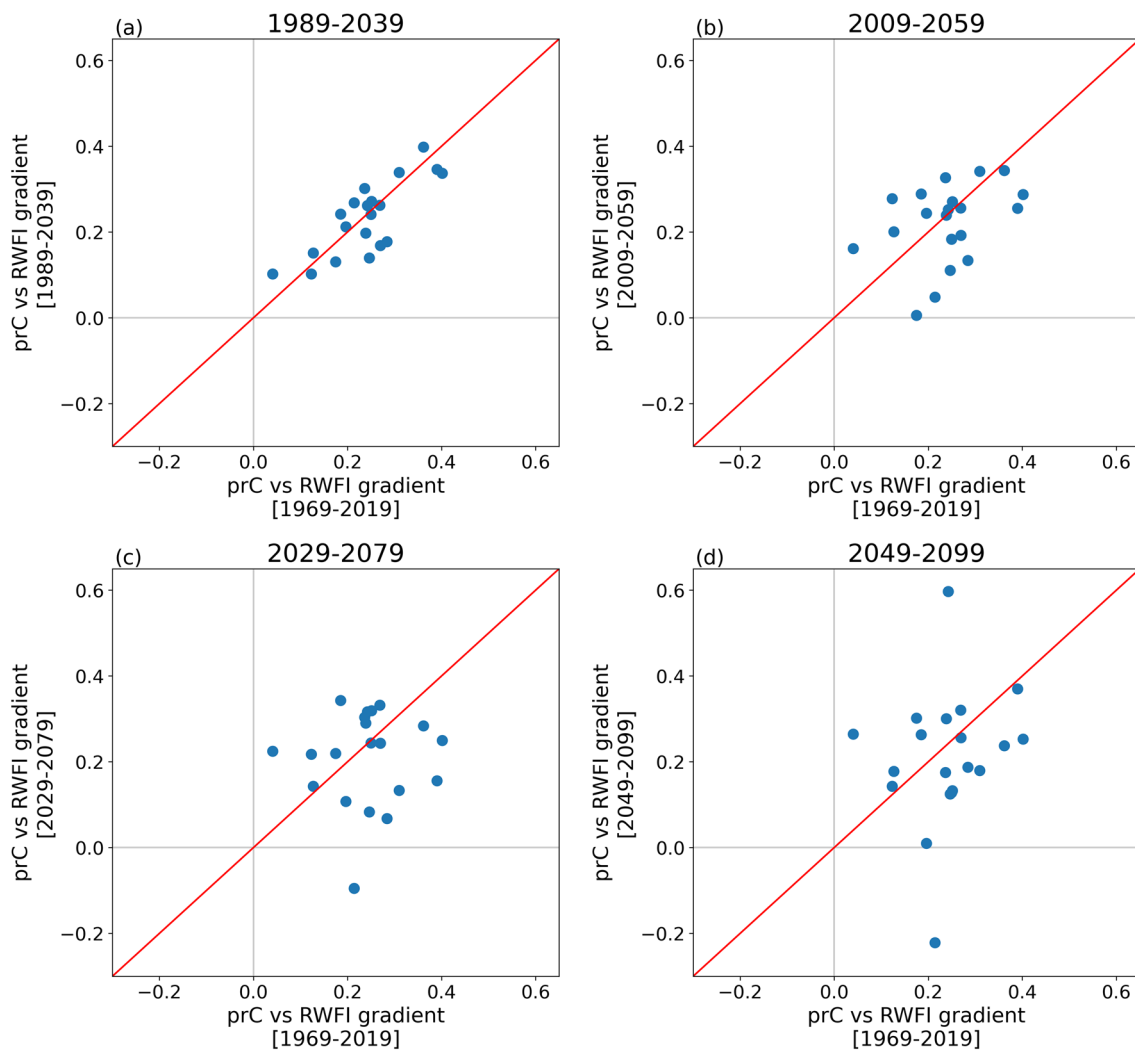


Fig. 11 Relationships between JJA mean precipitation in CE China (prC) and RWFI, compared for different time periods. Gradients of prC vs RWFI are shown for 4 50-year time periods: **a** 1989–2039, **b** 2009–2059, **c** 2029–2079 and **d** 2049–2099, and each are plotted against the gradient for 1969–2019. Each point represents a single

PPE member. The gradients were evaluated using detrended data, where 35-year running means were first subtracted from the prC and RWFI time series data. Detrending was applied separately for each PPE member

circulation, transient eddies, and a non-linear term (DMCD, DTE and DNL, respectively; see Eqs. 6–8 in Appendix B). The DMCD component, which describes moisture convergence changes resulting from changes in the mean circulation with the humidity held fixed (at present-day values), accounts for 24% of the variance in DP-DE (Fig. 13c). This relationship is clearly degraded compared to DDYN and suggests the other dynamical components (DTE, DNL and covariance terms) also contribute.

The contributions from all these components highlights the need for the in-depth moisture budget analysis over the simpler RWFI-based analysis we presented above. Like the DMCD component, the RWFI-based analysis attempted to capture the effect of changes in the mean circulation. However, they represent quite different ways to capture

the effects of the changing circulation: the DMCD component describes moisture convergence changes resulting from changes in the mean circulation with the humidity held fixed, whilst our RWFI-based analysis estimated the effects of holding the present-day relationship between precipitation and circulation anomalies fixed. Whilst we might expect some level of relationship between DMCD and DRWFI (indeed they are correlated with $R^2 = 0.53$), the latter clearly misses several aspects of the drivers of the precipitation change i.e., the remaining components of DDYN, as well as the contributions from DTH and DE. Figure 13d highlights this, which shows that DRWFI only explains a similar amount of the variance in DP-DE as DMCD.

Fig. 12 Changes in moisture budget components for CE China, based on JJA means for 2070–2099 vs 1980–2009. Values are shown for changes in precipitation (DP); precipitation minus evaporation (DP-DE); moisture convergence (DconvQ); the thermodynamic (DTH) and ‘dynamic’ contributions (DDYN); and a residual term (DRES) for the discrepancy between DP-DE and DTH+DDYN. The details of the calculation of these components are given in Appendix B. Each grey point represents a single PPE member. Examples from the discussion on sub-selection of the PPE (Sect. 6) are highlighted in colour

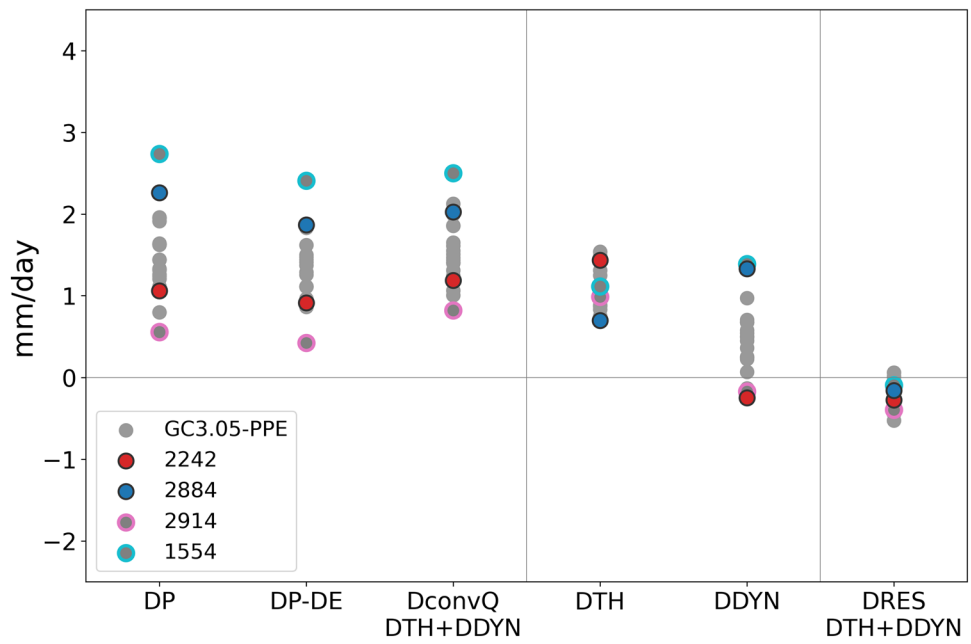
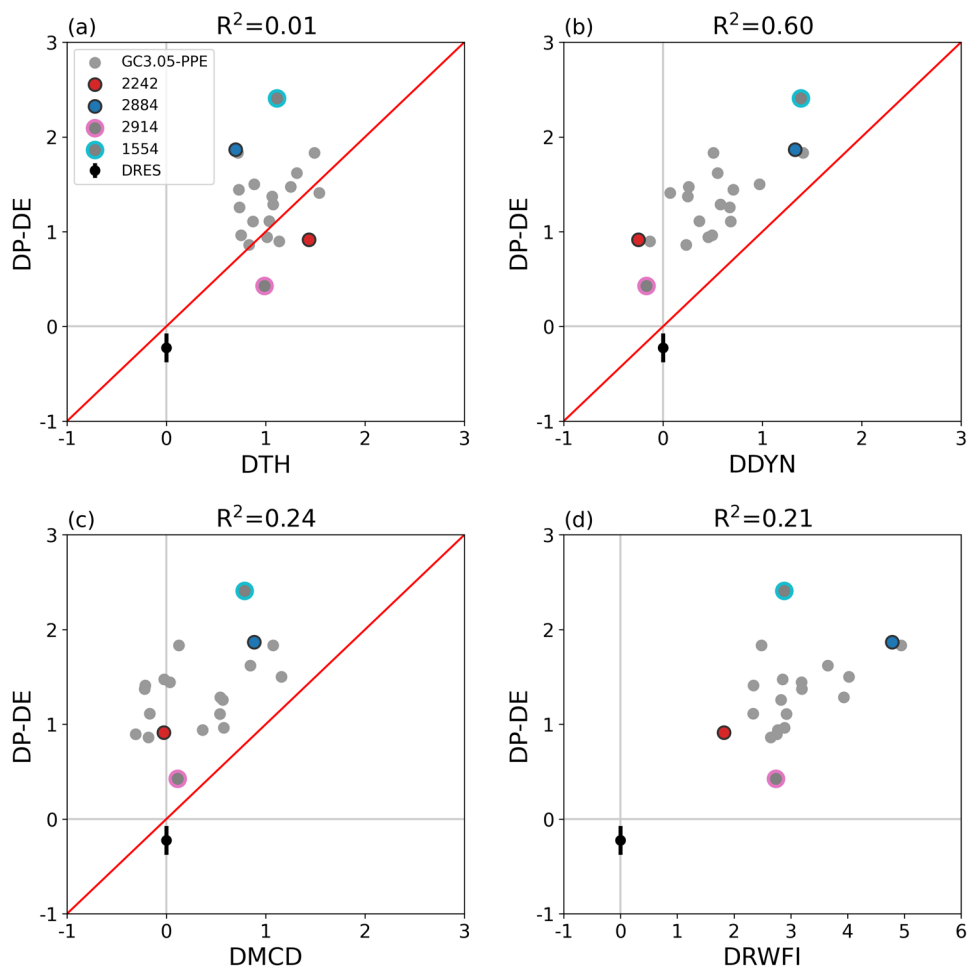


Fig. 13 Changes in precipitation minus evaporation (DP-DE) for CE China plotted against changes in: **a** the thermodynamic; **b** the ‘dynamic’; and **c** the mean-circulation dynamics components of the moisture budget for CE China. **d** DP-DE for CE China vs the RWFI. Changes are based on JJA means for 2070–2099 vs 1980–2009. The details of these calculations are given in Sect. 2.4.1 and Appendix B. The black point and error bar shows the mean and standard deviation of the residuals from the moisture budget analysis (DRES) to give an indication of the limit of confidence in the moisture budget component values. The remaining data are as described for Fig. 12. An estimate of the fraction of the variance in DP-DE explained is given in each case using the square of the Pearson correlation coefficient (R^2)



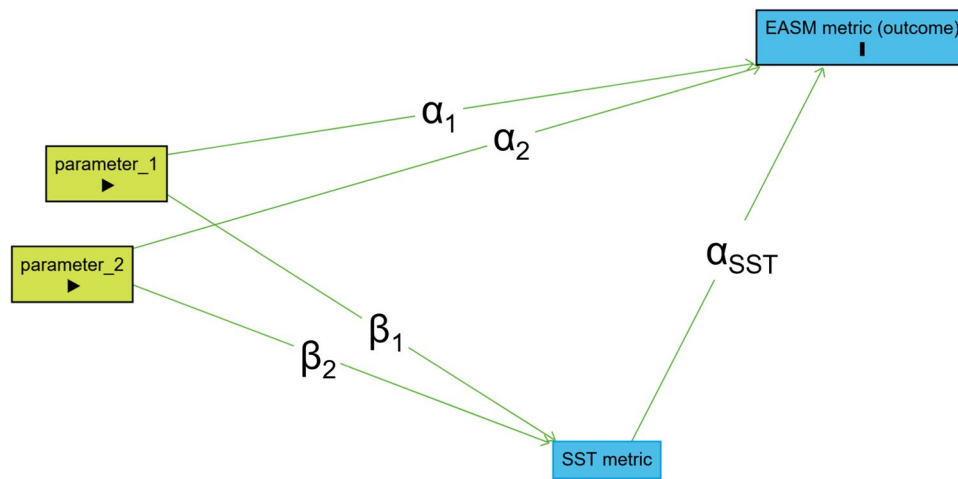


Fig. 14 Schematic of the causal network used to assess how PPE parameters affect metrics of the EASM in the GC3.05-PPE. Direct atmospheric effects are modelled for key parameters (denoted by arrows labelled α_1 and α_2) whilst controlling for the effect of changes in SSTs (α_{SST}). The indirect effects of the parameters i.e., those which

are mediated through changes in SSTs, are found by combining the impact of the parameters on the SSTs (arrows labelled β_1 and β_2) with that of the SSTs on the EASM (α_{SST}). See Appendix C for details. The schematic was built using Dagitty

Table 2 Results from the sensitivity analyses for the GA7.05-PPE and the linear regression analyses for the GC3.05-PPE, for each EASM metric

EASM metric	Parameter/SST metric	GA7.05-PPE sensitivity (% variance explained)	Linear regression coefficients		
			Direct effect	Indirect effect	Total effect
RWFI	ent_fac_dp	0.58	- 0.65	- 0.13	- 0.76
RWFI	amdet_fac	0.29	- 0.08	0.02	0.10
RWFI	ITC		0.32	-	0.32
Δ RWFI	ent_fac_dp	0.61	0.51	0.07	0.56
Δ RWFI	amdet_fac	0.25	0.09	0.00	- 0.04
Δ RWFI	Δ ITC		0.34	-	0.34

Sensitivity values for the GA7.05-PPE are given as the percentage of variance explained in the EASM metric by each parameter. For the GC3.05-PPE causal network regression analyses, the standardised regression coefficients are given, representing the effect of a change of 1 standard deviation of each predictor (the parameters or ITC metric) on the EASM metric

5 Sensitivity of the EASM to parameter perturbations

One advantage of the design of PPEs is that the difference between the configuration of each member is clear—it is the difference in the parameter values each member takes. This allows us to potentially trace different outcomes to specific parameters and processes within the model. Here, we present linear analyses of the relationships between parameters and the EASM in the GC3.05-PPE, using simple causal networks (Pearl 2009). The basic setup for these networks is shown in Fig. 14 (Textor et al. 2017; see Appendix C for details). We use these causal network to analyse the roles of ‘direct’ atmospheric effects, and ‘indirect’ effects which are mediated through the parameters’

influence on sea-surface temperature patterns (see Example 3, Kretschmer et al. 2021).

We could build and analyse causal networks for any of the EASM metrics we have studied in this paper. However, the precipitation metrics cover relatively small domains over China, and they can be influenced by many of the parameters/processes we perturb in the PPE. This can make the task of finding clear links between precipitation and model parameters challenging, especially given our limited sample size. Because of this we do not analyse parameter sensitivities for the precipitation metrics. We focus instead on the circulation metrics, namely present-day values and future changes in the reversed Wang and Fan index (RWFI and Δ RWFI).

One constraint on this analysis is the small number of members in the GC3.05-PPE (20) compared to the number of model parameters which are perturbed (47). This limits

our ability to detect links between the parameters and outcomes to the clearest cases, where only a few parameters have an influence. In light of this limitation, we use supporting evidence to help choose the parameters and SST metrics considered in each network, and to mitigate the rejection of potentially important effects (Amrhein et al. 2019).

To support the direct atmospheric effects, and to select the model parameters which potentially impact the EASM, we utilise a related PPE which uses the atmospheric component of HadGEM3-GC3.05 as its base model. We will refer to this as the GA7.05-PPE (Sexton et al. 2021). The GA7.05-PPE comprises around 500 members, which allows us to build emulators (statistical models that predict the effect of parameters on quantities of interest) and use these to perform more detailed sensitivity analyses of those quantities (Saltelli et al. 1999; Rostron et al. 2020). This is done for both the RWFI and Δ RWFI using the method of Rostron et al. (2020) to build each emulator. We use results from two 5-year fixed-SST experiments (Sexton et al. 2021): an amip-like experiment (2005–2009) to analyse the present-day RWFI values; and an amipFuture-like experiment, which uses a prescribed pattern of future SST warming (with a mean change of +4 K) for future changes in the RWFI. We use sensitivity analyses from the GA7.05-PPE to determine which model parameters individually explain more than 20% of the variance for each these metrics (Table 2). The selected model parameters are then used as the parameters for the corresponding causal network for the GC3.05-PPE.

The indirect effect in the causal networks is used to represent the impact of model parameters on SSTs, which then drive changes in the EASM. To represent the leading order effect of SSTs on the EASM we use a metric for the inter-hemispheric thermal contrast (ITC). Changes in the Asian monsoon has been found to vary in-phase with this SST pattern across a range of timescales, and future changes in the monsoon and the ITC have also been linked (Feudale and Kucharski 2013; Li et al. 2017; Chen et al. 2022). Here, we use the definition of the ITC from Chen et al. (2022), which is the difference in the area-averaged surface temperature between 20° N–50° N and 20° S–50° S.

The results from the causal networks are shown in Table 2. All the data is standardised before the coefficients in the network are determined using linear regression (see Appendix C). This means the coefficients represent the effect of a change of 1 standard deviation of each predictor (the parameters or ITC metric) on the outcome (EASM metric), and they can be compared directly.

For present-day values of the RWFI, parameters controlling the amplitude of deep-level entrainment (*ent_fac_dp*) and mixing detrainment (*amdet_fac*) are selected from the GA7.05-PPE sensitivity analysis. The coefficients from the GC3.05-PPE regression analysis suggests *ent_fac_dp* has a clear influence on the circulation and this is primarily a

direct effect (– 0.65 vs – 0.13). In contrast to the GA7.05-PPE, there is little detectable impact of *amdet_fac* for the GC3.05-PPE, for both the direct and indirect effects. However, there is a sizeable influence from SSTs (0.32), which may result from the effects of other parameters, or from internal variability.

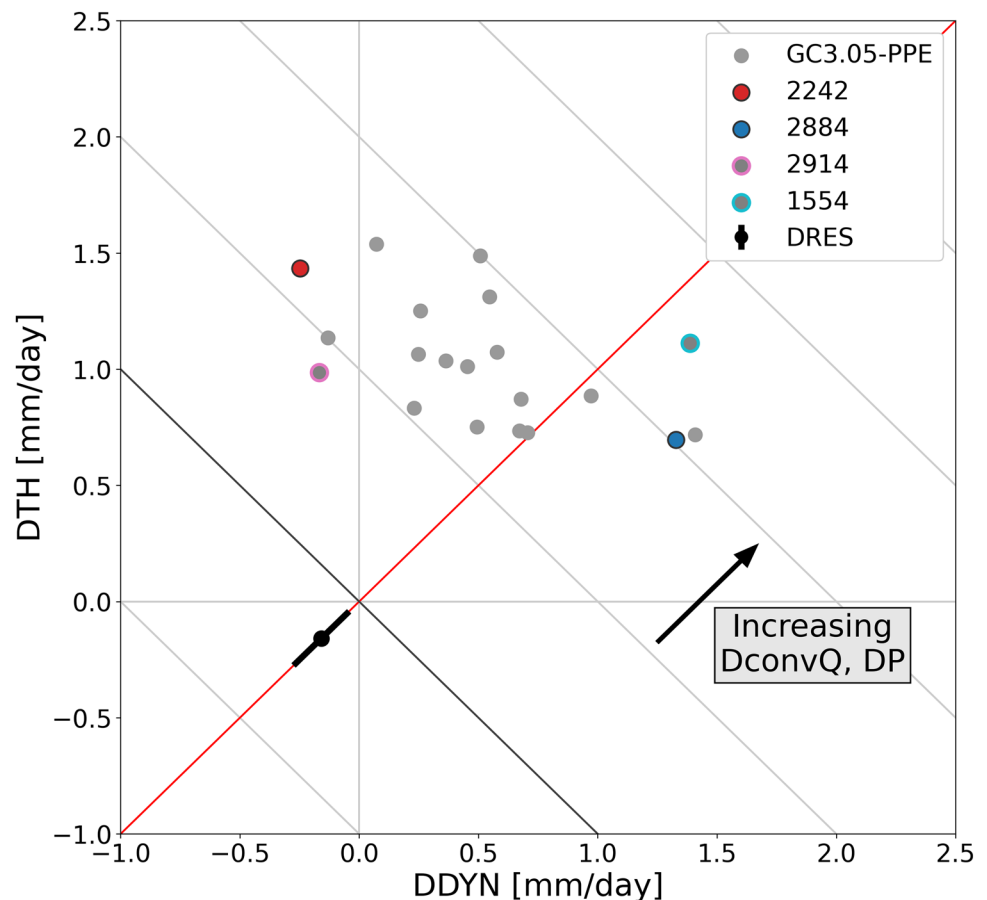
Increasing the value of *ent_fac_dp* leads to decreases in RWFI values i.e., it drives anomalous cyclonic circulation responses over the west Pacific region. This is true for both the GC3.05-PPE, as suggested by the negative regression coefficients in Table 2, and the GA7.05-PPE (not shown). These results are consistent with a previous study using an earlier configuration of the atmospheric component of the model (HadGEM3-GA3.0), where the Asian monsoon system's sensitivity to entrainment (and detrainment) parameters was tested (Bush et al. 2015). Figure 3 in that study shows a clear anomalous cyclonic circulation (which would be characterised by decreases in the RWFI) in response to increases in these parameters.

Studies based on more recent configurations of HadGEM3 have identified the Maritime Continent as a key region in the development of EASM circulation biases on sub-seasonal to seasonal timescales, which persist into climatological errors consistent with our results in Sect. 3.1 (Rodríguez and Milton 2019; Martin et al. 2021). One possible explanation for the sensitivity we have found to *ent_fac_dp* could be that an anomalous cyclonic circulation over the west Pacific (and an associated northward shift of the Hadley cell) results from a suppression of convective activity over the Maritime Continent, driven by increases in this parameter. Further work would be required to test this hypothesis in more detail e.g., by evaluating how *ent_fac_dp* values and circulation biases relate to metrics of ascending motion over the Maritime Continent.

For the future changes in RWFI (Δ RWFI) we again find that *ent_fac_dp* and *amdet_fac* are the key parameters in the GA7.05-PPE. Similarly to the sensitivities for present-day RWFI, *ent_fac_dp* has a clear direct impact on Δ RWFI in the GC3.05-PPE (with a regression coefficient of 0.51), but there is little evidence of an effect for *amdet_fac*. The SSTs again have an appreciable impact, with a regression coefficient of 0.34 associated with future changes in the ITC.

The positive coefficients for *ent_fac_dp* imply that members with higher values for this parameter have increasingly positive Δ RWFI. (A qualitatively similar relationship was also found for the GA7.05-PPE.) Since Δ RWFI is consistently positive in the GC3.05-PPE (Fig. 8a) this implies larger *ent_fac_dp* values drive future circulation changes over the west Pacific which are more anti-cyclonic. As discussed for the present-day RWFI sensitivity, one potentially key region here is the Maritime Continent. Determining how *ent_fac_dp* impacts future changes in convective activity over this region could be key to understanding the physical

Fig. 15 Changes in the thermodynamic (DTH) vs the dynamic (DDYN) components of the moisture budget for CE China (using JJA means for the periods 2070–2099 vs 1980–2009; see Appendix B for details). The data shown are as described for Fig. 13. Grey diagonals are lines of constant DTH+DDYN. Given small values for DRES and DE (see Fig. 12), these provides an estimate for changes in the moisture convergence (DconvQ) and precipitation (DP). The red line indicates where DTH and DDYN contribute equally to the moisture budget changes



mechanism linking this parameter to future changes in the EASM circulation in the HadGEM3-GC3.05 model. However, we will leave any further analysis of this to future work.

We have shown that changes across PPE members for both the RWFI and Δ RWFI are largely driven by the `ent_fac_dp` parameter. This shared dependence, with a negative sensitivity for RWFI and a positive sensitivity for Δ RWFI, is consistent with the negative correlation between these metrics shown in Fig. 9 (and described in Sect. 4.1).

We note that the sensitivities implied by the regression coefficients for the GC3.05-PPE are not always consistent with those from the GA7.05-PPE analysis. In every case the squared coefficients are smaller than the GA7.05-PPE explained variance. For `amdet_fac` in particular we found little evidence in the GC3.05-PPE for the sensitivities found in the GA7.05-PPE, for both the RWFI and Δ RWFI. These discrepancies can arise for several reasons. One may be the small number of members in the GC3.05-PPE. This can lead to large uncertainties on the regression coefficients when the GC3.05-PPE data is considered in isolation. However, this does not account for any supporting evidence e.g., the fact that these parameters were selected in the GA7.05-PPE sensitivity analysis, or qualitative agreement for the impact of the parameters between the GA7.05 and GC3.05-PPEs,

as was found for `ent_fac_dp`. Another source for the discrepancy may be from coupling effects: the atmosphere may respond differently in the presence of coupled SSTs compared to prescribed SSTs. There may also be non-linear effects, which would only be captured by the GA7.05-PPE sensitivity analysis.

6 Discussion on robustness and sub-selection of PPE members

The 20-member HadGEM3-GC3.05 PPE was developed to provide users with raw global model output, suitable for use in regional impacts and adaptation studies. However, a dataset of this size may not be practical or desirable to use, for example due to human or computational resource limitations. In such cases a subset of members could be chosen, which were deemed to be plausible but still representative of the future changes explored by the full ensemble (McSweeney et al. 2015). The details of this sub-selection will depend on the application, but here we discuss some considerations for sub-selecting the PPE in the context of the EASM.

A key consideration will be the robustness of the information provided by the models i.e., are they plausible? This could be based on the global and/or regional performance of the models. For example, Yamazaki et al. (2021) describe how members of the GC3.05-PPE were selected for inclusion in the UKCP18 project, based on: the performance of regional SSTs over the globe; the Atlantic Meridional Overturning Circulation (a key driver of climate and variability for the North Atlantic and Europe); twentieth century NH temperature trends; and present-day climatologies of temperature and precipitation over Europe. Using these, the PPE was reduced from 25 to 20 members (which we have used in this paper) in a first round of filtering, and to 15 members after a second round of filtering.

In this paper we have assessed leading-order metrics for the EASM, to give a broad indication of the performance of the PPE for this key mode of climate variability for China. For precipitation we have shown the PPE has a reasonable performance (compared to CMIP5 and CMIP6 models) in the mean states and variability. The performance does depend on the region (see Fig. 5), but the PPE has notably good performance for the key CE China region, covering the Yangtze River basin.

For low-level circulation we find a structural bias in the PPE, where all members are found to have a cyclonic bias in JJA over the East Asia/West Pacific region. In comparison, the CMIP models do encompass the observed mean state for the RWFI (Fig. 3a). However, since the CMIP5 and CMIP6 ensembles are comprised of single variants of different model structures, we cannot tell whether these models are themselves structurally biased. We note that several PPE models have a comparable level of performance to CMIP models for the RWFI, and that PPE models compare well in terms of RWFI variability (Fig. 3b). Of course, the tolerance level on the mean state could be set such that no PPE members are accepted, but this will depend on the user and could have implications for the level of diversity if set too strictly.

Once a plausible subset of models has been identified, users may want to sub-select further in a way that still captures the diversity in the projected changes being studied. As an example of this, we consider a simple sub-selection of the PPE for projections of CE China precipitation, based on the moisture budget analysis shown in Sect. 4.2.1. We have shown a range of changes for CE China precipitation across the PPE, and that changes in the dynamics (DDYN) are a key driver of this. In this context, a representative subset would target high and low values of changes in CE China precipitation and DDYN. We have highlighted two PPE members which do this in Figs. 12, 13, 14, 15, 16. Member 2242 (marked by a filled red circle) has a low precipitation change relative to other PPE members, driven in part by a slightly negative contribution from dynamical changes. In contrast, member 2884 (filled blue circle) has a relatively high change in CE China precipitation, with a large positive contribution from DDYN.

These two models sample high and low values in other metrics too, for example for changes in the mean circulation metrics (shown in Fig. 13c and d), and in their thermodynamic changes (Fig. 13a). The thermodynamic contributions of these two models partially offsets some of the differences from the dynamics i.e., DTH is large for in our low precipitation change scenario (member 2242), whilst DTH is small for our high precipitation change scenario (member 2884). This diversity is shown clearly in Figs. 15 and 16. In Fig. 15 our two example members lie in opposing corners of the DTH and DDYN values sampled by the full PPE and represent high and low precipitation changes due to the key role of the DDYN component. In Fig. 16 we show DTH fields overlaid with the mean 850 hPa wind field for 1980–2009 (top row) and DDYN fields overlaid with changes in the 850 hPa winds (2070–2099 vs 1980–2009; bottom row) for the two members. The 850 hPa winds are included to give an indication of the contribution of the mean circulation to both DTH and DDYN (see Eqs. 5 and 6 in Appendix B). The differences in the DDYN field, and the associated mean circulation changes, are striking and clearly affect the CE China region (covered by the red box), with large circulation changes bringing additional moisture into the region from the south for member 2884, but much weaker circulation changes for member 2242. Meanwhile, differences in DTH fields, and in the present-day circulation, are much more modest between the members – reflecting the smaller spread in DTH compared to DDYN.

Here we have covered one simple example of sub-selection, based on capturing diversity in the physical processes underlying twenty-first century precipitation changes. But there are many other ways a subset could be chosen. Even within the same framework of our moisture budget decomposition, other choices could be made. For example, we have highlighted two alternative (or additional) members in Figs. 12, 13, 14, 15 (see the cyan and pink circles), which also sample high and low DDYN values, although these members have very similar thermodynamic changes, in contrast to our earlier example. These members also have the highest/lowest CE China precipitation changes and could provide useful scenarios for studies of flooding and/or drought in the region. Alternative ways to sub-select might focus on other variables (e.g. temperatures or heat stress metrics for heat waves), or consider a wider set of metrics in multi-sector studies or more generic assessments (McSweeney et al. 2015; Palmer et al. 2021).

7 Summary

In this paper we have evaluated the simulation of the East Asian Summer Monsoon in a perturbed parameter ensemble of HadGEM3-GC3.05 coupled models by analysing their

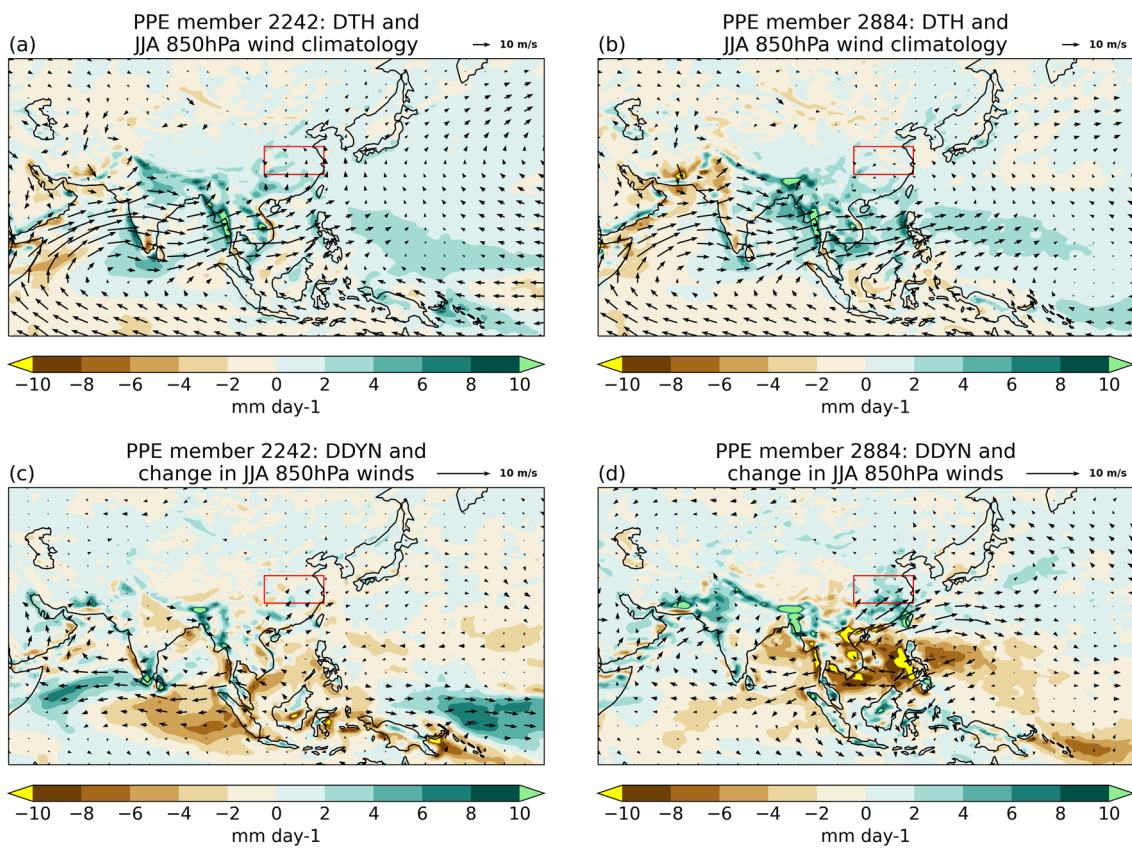


Fig. 16 **a** and **b** Maps of the changes in the thermodynamic (DTH) component of the moisture budget, overlaid with climatological 850 hPa winds, for members 2242 and 2884, respectively. **c** and **d** Changes in the dynamic (DDYN) component of the moisture budget

overlaid with changes in the 850 hPa winds for the same members. JJA means are shown in each case, with climatologies based on the period 1980–2009 and changes based on 2070–2099 vs 1980–2009. (See Appendix B for details of the moisture budget calculations.)

performance with respect to the observed climatology, and their projected changes. We focussed on leading-order metrics of the low-level (850 hPa) circulation and precipitation. In summary:

- For low-level circulation we find a structural bias in the PPE, where all members are found to have a cyclonic bias over the East Asia/West Pacific region (for JJA means). This bias is known to have affected previous Hadley Centre models, and the structural nature of the bias revealed by the PPE suggests it cannot be easily corrected by model parameter choices. Using the reversed Wang and Fan index (RWFI) we find negative biases are typical in the PPE and in CMIP5 and CMIP6 models, but there are examples of much smaller (and even positive) biases in CMIP5 and CMIP6.
- For precipitation we find the performance of PPE members varies spatially and temporally, with good performance for JJA climatologies in CE and NE China, but wet biases over southern China. The performance of CMIP models also varies by region, with differences

- between the performance of the PPE and CMIP models highlighting the benefits of considering both ensembles. Interannual variability is typically overestimated by both ensembles, but especially by the PPE. For seasonal cycles, we have indicated a split in the spatial and temporal modes of variability in the PPE, where members with smaller biases in southern China tend to show a seasonal cycle that peaks too early over CE China.
- The observed relationship between the RWFI and precipitation for CE China is captured in most PPE members. The PPE does not perform as well for the relationship between RWFI and Niño-3.4, with most members having a circulation response that is too weak.
- Changes for the twenty-first century for low-level circulation over the region are consistently anti-cyclonic in the PPE, suggesting a structural future change in the HadGEM3-GC3.05 model. There are examples of both cyclonic and anti-cyclonic circulation changes in CMIP5 and CMIP6, but without PPEs based on these models the structural nature of these changes is not known.

- Increases in mean precipitation are projected for the twenty-first century across China for all PPE members, and most CMIP members, with increases in variability also projected for all but a handful of members. For the CE China region, we find that both thermodynamic (atmospheric moistening) and dynamic (circulation) changes contribute to the increased precipitation, with the spread amongst PPE members largely caused by differences in the dynamic response.
- Using a parameter sensitivity analysis we found a parameter controlling the amplitude of deep-level entrainment is a key driver of the PPE spread for EASM circulation metrics.

We have also discussed how these projections may be used in practice, where considerations about the plausibility and usability of the models will be important, giving a simple example of sub-selecting PPE members aimed at capturing diversity in their projected precipitation changes. Users should also be aware of the limitations of these models in relation to structural biases which, as we have shown, are readily exposed by PPEs. Of course, these limitations apply to each of the CMIP5 and CMIP6 members but as we have highlighted through the text, one needs a PPE about each CMIP model to properly understand their individual structural biases. Indeed, as shown by Rostron et al. (2020), the errors of the tuned variant of HadGEM3-GA7.05 are not indicative of the structural model bias for all variables. Therefore, we encourage wider use of PPEs.

Appendix A

See Table 3.

Table 3 Sub-selected CMIP5 and CMIP6 models used in this study

	CMIP5 model	CMIP6 model
1	ACCESS1-3	ACCESS-CM2
2	BCC-CSM1.1	BCC-CSM2-MR
3	CCSM4	
4	CESM1-BGC	CESM2
5	CMCC-CM	CMCC-CM2-SR5
6	CNRM-CM5	CNRM-CM6-1
7	CanESM2	CanESM5
8	GFDL-ESM2G	GFDL-ESM4
9	HadGEM2-ES	HadGEM3-GC31-MM
10	IPSL-CM5A-MR	
11	MPI-ESM-MR	MPI-ESM1-2-HR
12	MRI-CGCM3	MRI-ESM2-0

Appendix B

In Sect. 4.2.1 we analyse twenty-first century changes in summer precipitation for the Central-East China region. Part of the analysis is based on changes in the components of the moisture budget for this region, which we calculate following a similar method to Seager et al. (2010). Following their Eq. 1, climatological averages of the difference between precipitation (P) and evapotranspiration (E) are balanced by the convergence of the vertically integrated moisture fluxes:

$$\bar{P} - \bar{E} = -\frac{1}{g} \int_0^{p_s} \nabla \cdot \overline{q\mathbf{v}} dp - q_s \mathbf{v}_s \cdot \nabla p_s + RES' \quad (1)$$

$$\bar{P} - \bar{E} = -\frac{1}{g} \int_0^{p_s} \nabla \cdot \left[\overline{q\mathbf{v}} + \overline{q'\mathbf{v}'} \right] dp + RES \quad (2)$$

here, p denotes pressure, q is the specific humidity and \mathbf{v} is the horizontal wind vector. Overbars denote monthly climatologies, while dashes denote the difference between daily mean values and the climatology (e.g., $q = \bar{q} + q'$). Surface values are denoted with subscript 's'. The term $q_s \mathbf{v}_s \cdot \nabla p_s$ in Eq. (1) accounts for the contribution to moisture fluxes from the flow at the surface (Seager et al. 2010). We incorporate this into a residual term (RES), which we use to account for imbalances between P-E and the vertically integrated moisture fluxes. Additional contributions to this residual will result from the fact the monthly means for the cross terms are derived from daily mean data (rather than timestep data), and from errors in the divergence and integration calculations (from the use of 17 discrete pressure levels), for example. From Eq. (1)–(2) we have used $\overline{q\mathbf{v}} = \overline{q\mathbf{v}} + \overline{q'\mathbf{v}'} + \overline{q'\mathbf{v}} + \overline{q\mathbf{v}'}$, where the cross terms in this equation are 0 (since $q' = \mathbf{v}' = 0$). At the timescales we consider in our analysis (30-year climatologies), changes in the total water content of the column are assumed to be negligible (Rodríguez et al. 2017).

We calculate changes in the moisture budget using differences in 30-year means, for the periods 2070–2099 vs 1980–2009. Denoting these differences as 'D', from Eq. (2):

$$D\bar{P} - D\bar{E} = -\frac{1}{g} \int_0^{p_s} \nabla \cdot \left[D(\overline{q\mathbf{v}}) + D(\overline{q'\mathbf{v}'}) \right] dp + DRES \quad (3)$$

$$D\bar{P} - D\bar{E} = -\frac{1}{g} \int_0^{p_s} \nabla \cdot \left[D(\overline{q\mathbf{v}}) + D(\overline{q'\mathbf{v}'}) \right] dp + DRES \quad (4)$$

We denote the change in the convergence of integrated moisture fluxes (i.e., the first term on the right-hand side of Eq. 4) as DconvQ. Further, using the notation from Seager et al. (2010), we define the contributions to DconvQ from changes in thermodynamics (TH), mean circulation dynamics (MCD), non-linear effects (NL) and transient eddies (TE) as:

$$DTH = -\frac{1}{g} \int_0^{p_s} \nabla \cdot (\bar{v} D\bar{q}) dp \tag{5}$$

$$DMCD = -\frac{1}{g} \int_0^{p_s} \nabla \cdot (\bar{q} D\bar{v}) dp \tag{6}$$

$$DNL = -\frac{1}{g} \int_0^{p_s} \nabla \cdot (D\bar{q} D\bar{v}) dp \tag{7}$$

$$DTE = -\frac{1}{g} \int_0^{p_s} \nabla \cdot D(\overline{q'v'}) dp \tag{8}$$

Conceptually, DTH is the change in moisture flux caused by changes in specific humidity carried on the present-day wind field, whilst DMCD is the contribution from changes in the wind field with specific humidity held to present-day values. DNL describes the effect of co-varying changes in the long-term averages of the humidity and the winds, whilst DTE is the contribution from changes in short-timescale co-variations in q and v .

Using these, we re-write Eq. (4) as:

$$D\bar{P} - D\bar{E} = D_{conv}Q + DRES \tag{9}$$

$$D\bar{P} - D\bar{E} = DTH + DMCD + DNL + DTE + DRES \tag{10}$$

Further, we define the sum of the contributions from the mean circulation dynamics, the transient eddies and the non-linear effects to be changes in the dynamics (DDYN):

$$DDYN = DMCD + DNL + DTE \tag{11}$$

So that:

$$D\bar{P} - D\bar{E} = DTH + DDYN + DRES \tag{12}$$

Our moisture budget calculations produce monthly means for each grid box. For the analysis presented in Sect. 4.2.1 we average these temporally and spatially to produce values for summer (JJA) over the Central-East China region (as defined in Sect. 2.3).

Appendix C

We use simple causal networks to describe the interactions between model parameters, sea-surface temperatures and EASM metrics in the GC3.05 PPE. The schematic in Fig. 14 shows an example of the network for 2 model parameters, although we could use more or fewer parameters in our analysis. The networks closely follow Example 3 from Kretschmer et al. (2021): a small number of parameters are modelled to have a ‘direct’ impact on the EASM (operating through the atmosphere), whilst also having an ‘indirect’ effect, which is mediated through differences in sea-surface temperatures (as measured by the ITC or Δ ITC).

The arrows in the causal network (Fig. 14) are important as they determine which terms need to be included in the multi-linear regression to obtain unbiased estimates of the causal effects, through the rules of causal inference theory (Pearl 2009). All data used in these regressions are standardised by first removing the mean, then normalising the residuals by the standard deviation. This allows the strength of the regression coefficients to be directly compared.

To measure the direct impact of the model parameters on the EASM i.e., that which does not involve changes in the SSTs, we need to block the indirect pathway by controlling for those SST changes. For the example in Fig. 14, the regression model is:

$$EASM = \alpha_1 * parameter_1 + \alpha_2 * parameter_2 + \alpha_{SST} * SST + \epsilon_D$$

The terms α_1 and α_2 are the direct effects of the parameters, whilst the α_{SST} term measures the influence of the SSTs on the EASM that is independent of the parameters. For the RWFI we use the ITC (as defined in Chen et al. 2022) as the SST metric, whilst for Δ RWFI we use Δ ITC.

To determine the indirect effect of the parameters on the EASM i.e., that which is mediated through changes in the SSTs, we first find the effect of the parameters on the SSTs separately using:

$$SST = \beta_1 * parameter_1 + \epsilon_{T1}$$

$$SST = \beta_2 * parameter_2 + \epsilon_{T2}$$

Multiplying these coefficients by α_{SST} (the strength of the SSTs effect on the EASM) gives the indirect effect for each parameter. We can also determine the total effect of each parameter on the EASM by regressing the EASM onto the parameters separately:

$$EASM = \gamma_1 * parameter_1 + \epsilon_{T1}$$

$$EASM = \gamma_2 * parameter_2 + \epsilon_{T2}$$

The total effects will approximately equal the sum of the direct and indirect effects, i.e.:

$$\gamma_1 \approx \alpha_1 + \beta_1 * \alpha_{SST}$$

$$\gamma_2 \approx \alpha_2 + \beta_2 * \alpha_{SST}$$

This ‘path-tracing’ rule (Pearl 2013; Kretschmer et al. 2021) arises because our analysis is based on linear regression, and because any correlations between parameter values is expected to be small.

Acknowledgements This work and its contributors (John W. Rostron, David M. H. Sexton, Kalli Furtado, Maria João Carvalho, Sean F. Milton, and José M. Rodríguez) were supported by the UK-China Research & Innovation Partnership Fund through the Met Office Climate Science

for Service Partnership (CSSP) China as part of the Newton Fund. Wenxia Zhang was supported by the National Natural Science Foundation of China (41905064).

Author contributions JWR, DMHS, KF, MJC and SFM contributed to the study conception and design. JWR and DMHS contributed the model data and performed the analyses. The first draft of the manuscript was written by JWR. All authors contributed to the interpretation of the results and provided comments and improvements to previous versions of the manuscript.

Funding This work and its contributors (John W. Rostron, David M. H. Sexton, Kalli Furtado, Maria João Carvalho, Sean F. Milton, and José M. Rodríguez) were supported by the UK-China Research & Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund (415000024199). Wenxia Zhang was supported by the National Natural Science Foundation of China (41905064).

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adler RF, Huffman GJ, Chang A et al (2003) The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J Hydrometeorol* 4:1147–1167. [https://doi.org/10.1175/1525-7541\(2003\)004%3c1147:TVGPCP%3e2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004%3c1147:TVGPCP%3e2.0.CO;2)
- Adler R, Sapiano M, Huffman G et al (2018) The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere* (basel) 9:138. <https://doi.org/10.3390/atmos9040138>
- Amrhein V, Greenland S, McShane B (2019) Scientists rise up against statistical significance. *Nature* 567:305–307. <https://doi.org/10.1038/d41586-019-00857-9>
- Bett PE, Scaife AA, Li C et al (2018) Seasonal forecasts of the summer 2016 Yangtze River Basin Rainfall. *Adv Atmos Sci* 35:918–926. <https://doi.org/10.1007/s00376-018-7210-y>
- Bush SJ, Turner AG, Woolnough SJ et al (2015) The effect of increased convective entrainment on Asian monsoon biases in the MetUM general circulation model. *Q J R Meteorol Soc* 141:311–326. <https://doi.org/10.1002/qj.2371>
- Chang C-P, Zhang Y, Li T (2000) Interannual and interdecadal variations of the East Asian Summer Monsoon and tropical Pacific SSTs. Part II: meridional structure of the monsoon. *J Clim* 13:4326–4340. [https://doi.org/10.1175/1520-0442\(2000\)013%3c4326:IAIVOT%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013%3c4326:IAIVOT%3e2.0.CO;2)
- Chen X, Wu P, Roberts MJ, Zhou T (2018) Potential underestimation of future Mei-Yu Rainfall with coarse-resolution climate models. *J Clim* 31:6711–6727. <https://doi.org/10.1175/JCLI-D-17-0741.1>
- Chen X, Zhou T, Wu P et al (2020a) Emergent constraints on future projections of the western North Pacific Subtropical High. *Nat Commun* 11:2802. <https://doi.org/10.1038/s41467-020-16631-9>
- Chen Z, Zhou T, Zhang L et al (2020b) Global land monsoon precipitation changes in CMIP6 projections. *Geophys Res Lett*. <https://doi.org/10.1029/2019GL086902>
- Chen Z, Zhou T, Chen X et al (2022) Observationally constrained projection of Afro-Asian monsoon precipitation. *Nat Commun* 13:2552. <https://doi.org/10.1038/s41467-022-30106-z>
- Collins M, Booth BBB, Bhaskaran B et al (2011) Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Clim Dyn* 36:1737–1766. <https://doi.org/10.1007/s00382-010-0808-0>
- Dee DP, Uppala SM, Simmons AJ et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597. <https://doi.org/10.1002/qj.828>
- Ding Y, Chan JCL (2005) The East Asian summer monsoon: an overview. *Meteorol Atmos Phys* 89:117–142. <https://doi.org/10.1007/s00703-005-0125-z>
- Feudale L, Kucharski F (2013) A common mode of variability of African and Indian monsoon rainfall at decadal timescale. *Clim Dyn* 41:243–254. <https://doi.org/10.1007/s00382-013-1827-4>
- Hardiman SC, Dunstone NJ, Scaife AA et al (2018) The asymmetric response of Yangtze river basin summer rainfall to El Niño/La Niña. *Environ Res Lett*. <https://doi.org/10.1088/1748-9326/aaa172>
- He C, Li T (2019) Does global warming amplify interannual climate variability? *Clim Dyn* 52:2667–2684. <https://doi.org/10.1007/s00382-018-4286-0>
- He C, Zhou T (2015) Responses of the Western North Pacific Subtropical high to global warming under RCP4.5 and RCP8.5 scenarios projected by 33 CMIP5 models: the dominance of tropical Indian Ocean-Tropical Western Pacific SST gradient. *J Clim* 28:365–380. <https://doi.org/10.1175/JCLI-D-13-00494.1>
- Hsu P, Li T, Luo J-J et al (2012) Increase of global monsoon area and precipitation under global warming: a robust signal? *Geophys Res Lett*. <https://doi.org/10.1029/2012GL051037>
- Hsu P, Li T, Murakami H, Kitoh A (2013) Future change of the global monsoon revealed from 19 CMIP5 models. *J Geophys Res Atmos* 118:1247–1260. <https://doi.org/10.1002/jgrd.50145>
- Huang X, Zhou T, Turner A et al (2020) The recent decline and recovery of Indian Summer Monsoon Rainfall: relative roles of external forcing and internal variability. *J Clim* 33:5035–5060. <https://doi.org/10.1175/JCLI-D-19-0833.1>
- Johnson JS, Regayre LA, Yoshioka M et al (2018) The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing. *Atmos Chem Phys* 18:13031–13053. <https://doi.org/10.5194/acp-18-13031-2018>
- Karmalkar AV, Sexton DMH, Murphy JM et al (2019) Finding plausible and diverse variants of a climate model. Part II: development and validation of methodology. *Clim Dyn* 53:847–877. <https://doi.org/10.1007/s00382-019-04617-3>
- Kitoh A, Endo H, Krishna Kumar K et al (2013) Monsoons in a changing world: a regional perspective in a global context. *J Geophys Res Atmos* 118:3053–3065. <https://doi.org/10.1002/jgrd.50258>

- Kretschmer M, Adams SV, Arribas A et al (2021) Quantifying causal pathways of teleconnections. *Bull Am Meteorol Soc* 102:E2247–E2263. <https://doi.org/10.1175/BAMS-D-20-0117.1>
- Li C, Scaife AA, Lu R et al (2016) Skillful seasonal prediction of Yangtze river valley summer rainfall. *Environ Res Lett*. <https://doi.org/10.1088/1748-9326/11/9/094002>
- Li Y, Ding Y, Li W (2017) Interdecadal variability of the Afro–Asian summer monsoon system. *Adv Atmos Sci* 34:833–846. <https://doi.org/10.1007/s00376-017-6247-7>
- Martin GM, Dunstone NJ, Scaife AA, Bett PE (2020) Predicting June mean rainfall in the Middle/Lower Yangtze River Basin. *Adv Atmos Sci* 37:29–41. <https://doi.org/10.1007/s00376-019-9051-8>
- Martin GM, Levine RC, Rodriguez JM, Vellinga M (2021) Understanding the development of systematic errors in the Asian summer monsoon. *Geosci Model Dev* 14:1007–1035. <https://doi.org/10.5194/gmd-14-1007-2021>
- McSweeney CF, Jones RG, Lee RW, Rowell DP (2015) Selecting CMIP5 GCMs for downscaling over multiple regions. *Clim Dyn* 44:3237–3260. <https://doi.org/10.1007/s00382-014-2418-8>
- Palmer TE, Booth BBB, McSweeney CF (2021) How does the CMIP6 ensemble change the picture for European climate projections? *Environ Res Lett* 16:094042. <https://doi.org/10.1088/1748-9326/ac1ed9>
- Pearl J (2009) Causal inference in statistics: an overview. *Stat Surv* 3:96–146. <https://doi.org/10.1214/09-SS057>
- Pearl J (2013) Linear models: a useful, “microscope” for causal analysis. *J Causal Inference* 1:155–170. <https://doi.org/10.1515/jci-2013-0003>
- Regayre LA, Johnson JS, Yoshioka M et al (2018) Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF. *Atmos Chem Phys* 18:9975–10006. <https://doi.org/10.5194/acp-18-9975-2018>
- Ringer MA, Martin GM, Greeves CZ et al (2006) The physical properties of the atmosphere in the New Hadley Centre Global Environmental Model (HadGEM1). Part II: aspects of variability and regional climate. *J Clim* 19:1302–1326. <https://doi.org/10.1175/JCLI3713.1>
- Rodríguez JM, Milton SF (2019) East Asian Summer atmospheric moisture transport and its response to interannual variability of the west Pacific subtropical high: an evaluation of the met office unified model. *Atmosphere (basel)* 10:457. <https://doi.org/10.3390/atmos10080457>
- Rodríguez JM, Milton SF, Marzin C (2017) The East Asian Atmospheric water cycle and monsoon circulation in the met office unified model. *J Geophys Res Atmos* 122:10246–10265. <https://doi.org/10.1002/2016JD025460>
- Ronghui H, Yifang W (1989) The influence of ENSO on the summer climate change in China and its mechanism. *Adv Atmos Sci* 6:21–32. <https://doi.org/10.1007/BF02656915>
- Rostron JW, Sexton DMH, McSweeney CF et al (2020) The impact of performance filtering on climate feedbacks in a perturbed parameter ensemble. *Clim Dyn*. <https://doi.org/10.1007/s00382-020-05281-8>
- Saltelli A, Tarantola S, Chan KP-S (1999) A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41:39–56. <https://doi.org/10.1080/00401706.1999.10485594>
- Sanderson BM (2011) A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. *J Clim* 24:1362–1377. <https://doi.org/10.1175/2010JCLI3498.1>
- Seager R, Naik N, Vecchi GA (2010) Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming. *J Clim* 23:4651–4668. <https://doi.org/10.1175/2010JCLI3655.1>
- Sexton DMH, Karmalkar AV, Murphy JM et al (2019) Finding plausible and diverse variants of a climate model. Part 1: establishing the relationship between errors at weather and climate time scales. *Clim Dyn* 53:989–1022. <https://doi.org/10.1007/s00382-019-04625-3>
- Sexton DMH, McSweeney CF, Rostron JW et al (2021) A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 1: selecting the parameter combinations. *Clim Dyn*. <https://doi.org/10.1007/s00382-021-05709-9>
- Song F, Zhou T (2014a) The climatology and interannual variability of East Asian Summer Monsoon in CMIP5 coupled models: does air-sea coupling improve the simulations? *J Clim* 27:8761–8777. <https://doi.org/10.1175/JCLI-D-14-00396.1>
- Song F, Zhou T (2014b) Interannual variability of East Asian Summer Monsoon simulated by CMIP3 and CMIP5 AGCMs: skill dependence on Indian Ocean–Western Pacific Anticyclone Teleconnection. *J Clim* 27:1679–1697. <https://doi.org/10.1175/JCLI-D-13-00248.1>
- Sperber KR, Annamalai H, Kang I-S et al (2013) The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim Dyn* 41:2711–2744. <https://doi.org/10.1007/s00382-012-1607-6>
- Textor J, van der Zander B, Gilthorpe MS et al (2017) Robust causal inference using directed acyclic graphs: the R package ‘dagitty.’ *Int J Epidemiol* 45:dyw341. <https://doi.org/10.1093/ije/dyw341>
- Tsushima Y, Ringer MA, Martin GM et al (2020) Investigating physical constraints on climate feedbacks using a perturbed parameter ensemble. *Clim Dyn*. <https://doi.org/10.1007/s00382-020-05318-y>
- Wagman BM, Jackson CS (2018) A test of emergent constraints on cloud feedback and climate sensitivity using a calibrated single-model ensemble. *J Clim* 31:7515–7532. <https://doi.org/10.1175/JCLI-D-17-0682.1>
- Walters D, Boutle I, Brooks M et al (2017) The met office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geosci Model Dev* 10:1487–1520. <https://doi.org/10.5194/gmd-10-1487-2017>
- Wang B, Fan Z (1999) Choice of South Asian Summer Monsoon Indices. *Bull Am Meteorol Soc* 80:629–638. [https://doi.org/10.1175/1520-0477\(1999\)080%3c0629:COASM%3e2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080%3c0629:COASM%3e2.0.CO;2)
- Wang B, Wu R, Fu X (2000) Pacific–East Asian teleconnection: how does ENSO affect East Asian Climate? *J Clim* 13:1517–1536. [https://doi.org/10.1175/1520-0442\(2000\)013%3c1517:PEATHD%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013%3c1517:PEATHD%3e2.0.CO;2)
- Wang B, Wu Z, Li J et al (2008) How to measure the strength of the East Asian Summer Monsoon. *J Clim* 21:4449–4463. <https://doi.org/10.1175/2008JCLI2183.1>
- Wang H, Fan K, Sun J et al (2015) A review of seasonal climate prediction research in China. *Adv Atmos Sci* 32:149–168. <https://doi.org/10.1007/s00376-014-0016-7>
- Wang PX, Wang B, Cheng H et al (2017) The global monsoon across time scales: mechanisms and outstanding issues. *Earth-Sci Rev* 174:84–121. <https://doi.org/10.1016/j.earscirev.2017.07.006>
- Williams KD, Copsey D, Blockley EW et al (2018) The met office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *J Adv Model Earth Syst* 10:357–380. <https://doi.org/10.1002/2017MS001115>
- Wu R, Hu Z-Z, Kirtman BP (2003) Evolution of ENSO-related rainfall anomalies in East Asia. *J Clim* 16:3742–3758. [https://doi.org/10.1175/1520-0442\(2003\)016%3c3742:EOERAI%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016%3c3742:EOERAI%3e2.0.CO;2)
- Xie SP, Zhou ZQ (2017) Seasonal modulations of El Niño-related atmospheric variability: Indo–Western Pacific Ocean feedback. *J Clim* 30:3461–3472. <https://doi.org/10.1175/JCLI-D-16-0713.1>
- Xie S-P, Kosaka Y, Du Y et al (2016) Indo-western Pacific ocean capacitor and coherent climate anomalies in post-ENSO

- summer: a review. *Adv Atmos Sci* 33:411–432. <https://doi.org/10.1007/s00376-015-5192-6>
- Yamazaki K, Sexton DMH, Rostron JW et al (2021) A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 2: global performance and future changes. *Clim Dyn*. <https://doi.org/10.1007/s00382-020-05608-5>
- Yang B, Zhang Y, Qian Y et al (2015) Parametric sensitivity analysis for the Asian Summer Monsoon precipitation simulation in the Beijing Climate Center AGCM, version 2.1. *J Clim* 28:5622–5644. <https://doi.org/10.1175/JCLI-D-14-00655.1>
- Yokohata T, Annan JD, Collins M et al (2013) Reliability and importance of structural diversity of climate model ensembles. *Clim Dyn* 41:2745–2763. <https://doi.org/10.1007/s00382-013-1733-9>
- Zhang W, Furtado K, Wu P et al (2021) Increasing precipitation variability on daily-to-multiyear time scales in a warmer world. *Sci Adv* 7:1–12. <https://doi.org/10.1126/sciadv.abf8021>
- Zhou T-J (2005) Atmospheric water vapor transport associated with typical anomalous summer rainfall patterns in China. *J Geophys Res* 110:D08104. <https://doi.org/10.1029/2004JD005413>
- Zhou S, Huang G, Huang P (2018) Changes in the East Asian summer monsoon rainfall under global warming: moisture budget decompositions and the sources of uncertainty. *Clim Dyn* 51:1363–1373. <https://doi.org/10.1007/s00382-017-3959-4>
- Zhou T, Lu J, Zhang W, Chen Z (2020) The Sources of Uncertainty in the Projection of Global Land Monsoon Precipitation. *Geophys Res Lett*. <https://doi.org/10.1029/2020GL088415>
- Murphy JM, Harris GR, Sexton DMH et al (2018) UKCP18 land projections: science report
- Wilks DS (2011) *Statistical methods in the atmospheric sciences*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.